

GenCore version 5.1.6  
Copyright (c) 1993 - 2004, Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:11:40 ; Search time 44 Seconds  
(without alignments)  
1016.836 Million cell updates/sec

Title: US-09-942-146A-1  
Perfect score: 2451  
Sequence: 1 MGRKEMVDRVPKMFVLISI.....PEHRAVNLSNSLWMLQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*  
1: piri:\*  
2: piri2:\*  
3: piri3:\*  
4: piri4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2416.5	98.6	466	2 S09837	hypothetical prote
2	122	5.0	475	2 G72274	glutaryl tRNA-Gln
3	116.5	4.8	1802	2 S69703	HKE1 protein precu
4	111	4.5	1203	2 T31029	related to pathway
5	110.5	4.5	739	2 T21769	hypothetical prote
6	109	4.4	651	2 C56653	membrane glycoprot
7	108.5	4.4	1224	2 E71611	hypothetical prote
8	107.5	4.4	919	2 T16693	hypothetical prote
9	106	4.3	1285	2 S70582	botulinum neurotox
10	103.5	4.2	2101	2 S7245	insulin receptor (
11	103	4.2	2028	2 T08025	DNA-directed RNA p
12	102.5	4.2	505	2 B46629	mucin 6, gastric (
13	102.5	4.2	638	2 T38863	hypothetical prote
14	101.5	4.1	1169	2 S38181	flocculation prote
15	101.5	4.1	1848	2 A44140	cellulose-binding
16	101	4.1	880	2 S60137	beta-N-acetylhexos
17	100.5	4.1	1995	2 T08166	probable membrane
18	100.5	4.1	3570	2 T40065	mucin MUC5B, trach
19	100	4.1	660	2 JW0067	chitinase (EC 3.2.
20	100	4.1	1429	2 T41699	C2-domain family p
21	99.5	4.1	369	2 C88030	protein F46P5.10 (
22	99.5	4.1	1348	2 AHI115	cell surface prote
23	99	4.0	375	2 H70226	conserved hypotet
24	99	4.0	621	2 C81026	ABC transporter, A
25	98.5	4.0	738	2 T44194	hypothetical prote
26	98.5	4.0	738	2 T44007	hypothetical prote
27	98.5	4.0	1365	2 T18419	hypothetical prote
28	97.5	4.0	1174	1 HUBYDH	helicase (EC 3.6.1
29	97	4.0	452	2 B97095	membrane associate

30	97	4.0	621	2 H81971	probable ABC trans
31	97	4.0	1777	2 T34369	hypothetical prote
32	96.5	3.9	347	1 DNHUN2	NADH2 dehydrogenas
33	96.5	3.9	3890	2 C89921	hypothetical prote
34	96	3.9	721	2 B90487	hypothetical prote
35	96	3.9	722	2 T00049	hemocyte protein A
36	96	3.9	809	2 T39626	hypothetical prote
37	96	3.9	1234	2 B36186	I factor protein 2
38	96	3.9	2437	2 T18482	hypothetical prote
39	95.5	3.9	474	2 B75126	hypothetical prote
40	95.5	3.9	653	2 H82316	conserved hypotet
41	95	3.9	291	2 A99181	xerC/D integrase-r
42	95	3.9	522	2 T37536	cell polarity prot
43	95	3.9	534	2 T33393	hypothetical prote
44	95	3.9	796	2 T21460	hypothetical prote
45	95	3.9	899	2 C71608	origin recognition

ALIGNMENTS

RESULT 1

S09837

hypothetical protein UL74 precursor - human cytomegalovirus (strain AD169)  
C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004

C:Accession: S09837

R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.;

M.J. Barrell, B.G.

Cur. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis-of-the-protein-coding-content-of the sequence of human cytomegalovirus

A:Reference number: S09749; MUID:90269039; PMID:2161319

A:Accession: S09837

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-466 <CHE>

A:Cross-references: UNIPROT:P16750; EMBL:X17403; NID:G59591; PIDN:CAA35389.1; PID:G17808

A:Note: this sequence was submitted to the EMBL Data Library, December 1989

C:Superfamily: human cytomegalovirus hypothetical protein UL74

C:Keywords: glycoprotein

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-466/Product: hypothetical protein UL74 #status predicted <MAT>

F:75,83,87,103,130,157,162,171,219,242,288,292,350,367,385,392,399,433,443,454/Binding s

Query Match 98.6%; Score 2416.5; DB 2; Length 466;

Best Local Similarity 98.9%; Pred. No. 4.5e-175;

Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 MGRKEMVDRVPKMFVLISISFLVSPFINKVMSKALYNRPWGLVLSKIGYKLDQLKL 60

Db 1 MGRKEMVDRVPKMFVLISISFLVSPFINKVMSKALYNRPWGLVLSKIGYKLDQLKL 60

Qy 61 EILRQLETTTSTKYNVSKQPVKNLTWKTTPPPQYVILAGPIQNTSYITVWDFDFSTQLRK 119

Db 61 EILRQLETTTSTKYNVSKQPVKNLTWKTTPPPQYVILAGPIQNTSYITVWDFDFSTQLRK 120

Qy 120 PAKYVYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179

Db 121 PAKYVYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180

Qy 180 NVPRWNTKLYVGPVKVNDVSGTIVFLGLTALLRAYAQRNCTHSPFLVNMARNLFRVPKY 239

Db 181 NVPRWNTKLYVGPVKVNDVSGTIVFLGLTALLRAYAQRNCTHSPFLVNMARNLFRVPKY 240

Qy 240 INGTKLNTWRKLKRQAPVKEQEKAKKTQSTTTTPYFSYTTSAALNVTNTVYSITTA 299

Db 241 INGTKLNTWRKLKRQAPVKEQEKAKKTQSTTTTPYFSYTTSAALNVTNTVYSITTA 300

Qy 300 ARRVSTSTIAYRPDSSFMKSIATQLRDLATVWYTTTLRYRONPCEPSRNRTAVSEFMKN 359

Db 301 ARRVSTSTIAYRPDSSFMKSIATQLRDLATVWYTTTLRYRONPCEPSRNRTAVSEFMKN 360

Qy 360 THVLRNETPTIYGTILDMSSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWD 419  
Db 361 THVLRNETPTIYGTILDMSSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWD 420  
Qy 420 YLDSLLFLDEIRNFSLSRSPPTVNLTPPEHRAVNLSTNSLWMLQ 465  
Db 421 YLDSLLFLDEIRNFSLSRSPPTVNLTPPEHRAVNLSTNSLWMLQ 466

RESULT 2  
glutamyl tRNA-Gln amidotransferase, subunit A - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: G72274  
R:Nelson, K.B.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.

Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: G72274  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-475 <ARN>  
A:Cross-references: UNIPROT:Q9X0Z9; GB:AE001782; GB:AE000512; NID:g4981821; PIDN:AAD3634  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1272  
C:Superfamily: indoleacetamide hydrolase

Query Match 5.0%; Score 122; DB 2; Length 475;  
Best Local Similarity 21.2%; Pred. No. 0.16;  
Matches 101; Conservative 64; Mismatches 171; Indels 140; Gaps 23;

Qy 17 LISISFLLYSFINKVMSKALYNRPWRLGLVLSKIGKYLKIDQLKRLRLQLE----TTTST 72  
Db 1 MIDLDPRKUTIEC-----LKLSEEREKLPQLSLETIKELDPHVKAFTSV 46  
Qy 73 KNVVS-----KQPVKNLTMTNTEFPQYILAGPIQNSITYLWFDYFSTQLR 118  
Db 47 RENVSEKKGKFWGIPVAIKONILTLGMRIT-----CASRIENYESV---FDATVVKKM 98  
Qy 119 KPAKYV-----YSQYNHTAKITP---RPPCGRVP-----SMTCLSE-MLNVSK 159  
Db 99 KEAGFVVVGKANLDEFAMGSSSTERSAFTTRNPWDLERVPGSSGSGSAAAVSAGVVAAL 158  
Qy 160 RNDTG-----EQCGNFTTFNPMFFNVPWNNTKLY-----VGP-TKVNVDSTIYFLG 206  
Db 159 GSDTGGSVRQPASLQGVGKFTYGLVSRYLVAFASSLDQIGPITKTVRDA----- 210  
Qy 207 LTALLLRYAQRNCTHSYLVNMSRLMFRVPKYINGTKLKNTRKLRK-KQAPVKEQPEK 265  
Db 211 --ATLMEIISGRDNDATTNRRKVDPLSEIEEGVSGMKFAVPEIIEHDIIEGVSRFEE 268  
Qy 266 KAK-----KTQSTTPPYFSYTTSAALNVTTNVTYSITTAARRVSTSTIARPDSSFMK 318  
Db 269 AKLLERLGAQVERKLPKIKYSA-----TYVVIAPAE--ASSNLARPDGVKYL 317  
Qy 319 SIMATQLRLDATWYVTLRLVRQNPFCPSRNRTAVSEFMKNVTLRNPTIYIYGLDM 378  
Db 318 RIKEKGLREM-----YMKTRNVGFGSEVRRIMI-----GTFTL 351  
Qy 379 SSLY-----NETMFVENKTASDSNKT-----TPTSPSMGQRTFI-DPLWDYL 421  
Db 352 SAAYEAYFNKAMKVRKRKISDELNEVLQYDAILTPISPTAFKIGIKDPLTYVL 407

RESULT 3  
HKR1 protein precursor - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YDR420W  
C:Species: Saccharomyces cerevisiae

C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 09-Jul-2004  
C:Accession: S69703; A53382  
R:Dietrich, F.S.  
A:Description: The sequence of S. cerevisiae lambda 3641 and cosmids 9461, 9831, and 9411  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S69555  
A:Accession: S69703  
A:Molecule type: DNA  
A:Residues: 1-1802 <DIE>  
A:Cross-references: UNIPROT:Q04051; EMBL:U33007; NID:9327685; PIDN:AAB64857.1; PID:929276;  
R:Kasahara, S.; Yamada, H.; Mio, T.; Shiratori, Y.; Miyamoto, C.; Yabe, T.; Nakajima, T.  
J. Bacteriol. 176, 1488-1499, 1994  
A:Title: Cloning of the Saccharomyces cerevisiae gene whose overexpression overcomes the  
A:Reference number: A53382; MUID:94156857; PMID:8113191  
A:Accession: A53382  
A:Molecule type: DNA  
A:Residues: 1-581, 'A', 593-593, 'A', 595-1802 <KAS>  
A:Cross-references: EMBL:S69101; NID:9545659; PIDN:AAB30051.1; PID:9545660  
A:Experimental source: YNN295  
A:Note: sequence extracted from NCBI backbone (NCBIN:144410, NCBI:P:144411)  
C:Genetics:  
A:Gene: SGD:HKR1  
A:Cross-references: SGD:S0002828; MIPS:YDR420W  
A:Map position: 4R  
C:Keywords: calcium binding; glycoprotein; transmembrane protein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-1802/Product: HKR1 protein #status predicted <MAT>  
F:1483-1508/Domain: transmembrane #status predicted <TM>  
F:1645-1656/Domain: calcium binding #status predicted <CAL>

Query Match 4.8%; Score 116.5; DB 2; Length 1802;  
Best Local Similarity 20.7%; Pred. No. 2.5; Indels 127; Gaps 19;  
Matches 90; Conservative 60; Mismatches 157;

Qy 68 TTISTKYNKQPVKNLTMTNTEFPQYILAGPIQNSITYLWFDYFSTQLRKPAKYVYSL 127  
Db 783 TSTSSPDYVYSTFAAIISSGYT-----PSPASVA-----MSTSSSSPDYVYSL 830  
Qy 128 YNHTAK-----TIFRPPPCGRVPSMTCLSEMLNVKRNKTGEQCGNFTTENPMF---FN 180  
Db 831 SSGASRSSIATYFSPSPSTSLPTSS-----TYTFSSAYAFSPS 870  
Qy 181 VPRNNTKLYVPTKVNVD-SQTIYFLGLTALLR--YAQRNCTHS-----FYLVN 227  
Db 971 SERYSTSTTAPQIHSTLRSITDFLLQTSMAIQSVSQISTSTLNDEIHSSALLSVFN 930  
Qy 228 AMSNLFPRVPKYINGTKLKNTRKLRKQAPVKEQPEKAKKT-----QSTTT 275  
Db 931 PSASNLVETSLIISSTQASITSPKNSAKISSLQSLSSSTKNPYDTANKNTSGRSTVV 990  
Qy 276 PYFSYTTSAALNVTTNVTYSIT-TAARRVSTSTIAY-----RP 312  
Db 991 SNFLYTSAA--KPDNEKFSATPTEITTISSSHAYSLSPSSHNSVTGLSHNFPVDSKS 1048  
Qy 313 DSSF-----MKSIMATQLRLDATWYVTLRLVRQNPFCPSERNRTAV 353  
Db 1049 ATSGYSSSSSISSIKLSKETIPASKSVSNITQER--ITSFTSTL--RANSQSEKSEGRNSV 1104  
Qy 354 SEFMKNVTLRNPTIYIYGT-LDMSLY-----NETMFVENKTASDSNKTTP 402  
Db 1105 GS-LQSSHI---SSNPSTLNTKVDKSLSKRVKMTWGENGEETGLTTTKYKSSSETS 1160  
Qy 403 TSPSMGQRTFIDP 416  
Db 1161 GSYRSFTKISIGP 1174

RESULT 4  
TS1029  
related to pathway-specific nitrogen regulator [imported] - Neurospora crassa  
N:Alternate names: protein B7F21.100  
C:Species: Neurospora crassa  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 09-Jul-2004

A;Gene: CBSP:F35B12.7  
A;Map position: 5  
A;Introns: 26/1; 61/3; 92/2; 131/1; 164/3; 240/1; 295/3; 319/2; 428/2; 482/3; 704/3

Query Match      4.5%; Score 110.5; DB 2; Length 739;  
Best Local Similarity 21.3%; Pred.No. 2.1;  
Matches 83; Conservative 42; Mismatches 151; Indels 113; Gaps 17;

Qy		79 QPV-----KNLTNTEPPQQYILAGP-----IQNSITYL-----WPDF-----YY 113     :       :
Db		313 QPVLKYPYSTNSSFPQ--LIPSPMATPHIHNASVTFVVVKPGVPQNDFAAAPQSRSVY 370     :       :
Qy		114 STQLRKPA-----KYVISQ----YNHTAKTITPR-----PPGCRVPSTMCLSEMLN 156     :       :
Db		371 SPOIWPAAEPKEDYTFSDPKVYNFSINLQT VKLENDGDDELVDVGVSVDGLTLTDKKYN 430     :       :
Qy		157 VSKRNTDGECCGNFTTFNPMPFNVRWNTKL-----YVGPTKANVDSOTIIFLGIGITAL 210     :       :
Db		431 KDKSADAYLSGIQR-----YLNLKYTSKTSEVINFMIDRLMDPGTT----- 474     :       :
Qy		211 LLRYAORNCSTHPSLYLVNAMSRNFLFRYPKVINGTKLNMTMKLRKOAPVKEQEKAOKAT 270     :       :
Db		475 -----QPGRIVTVSGSFST-----APIVTSVSTRATR 503     :       :
Qy		271 QSITTPFYSVTTSAALN---VTNN-----VTYISITTAARVSVSTAIARP---DSFPMKS 319     :       :
Db		504 SSSTAPVSVTVTITSQSIVTTNKSSPITSXTTTSINVSSSTKTRMQSTTTYQKL 563     :       :
Qy		320 IMATQLRDLATWYTILRVQRNFPCPSNRRTAVSFFMKNTHVLIRNETPTYTYIGILDMS 379     :       :
Db		564 AQASASTASTPIBITTONISTASIAPS-TKTPITPSLIITRSSTTKLSDIVI-----TKS 618     :       :
Qy		380 SLYYNETMFVENKTASDNKTTPPTSMSG 408     :       :
Db		619 SLAYTATPTSAPTSAKITTHITTPIP 647     :       :

RESULT 6  
C56653  
membrane glycoprotein precursor - human herpesvirus 6 (strain UI102)  
C;Species: human herpesvirus 6  
C;Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 09-Jul-2004  
C;Accession: C56653  
R;Compelt, U.A.; Carss, A.L.; Sun, N.; Arrand, J.R.  
DNA Seq. 3, 25-39, 1992  
A;Title: Infectivity determinants encoded in a conserved gene block of human herpesvirus 6  
A;Reference number: A56653; UID:93091236; PMID:1333836  
A;Accession: C56653  
A>Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-651 <DOM>  
A;Cross-references: UNIPROT:Q06093; GB:X83413; MID:g853961; PIDN:CAA58381.1; PID:g8540202  
A;Experimental source: strain UI102  
A>Note: sequence extracted from NCBI backbone (NCBIN:120533, NCIBIP:120546)  
C;Superfamily: herpesvirus glycoprotein H  
C;Keywords: glycoprotein

Query Match      4.4%; Score 109; DB 2; Length 651;  
Best Local Similarity 18.9%; Pred.No. 2.3;  
Matches 107; Conservative 77; Mismatches 189; Indels 192; Gaps 25;

Qy		5 EMNVDRDPKFVLISIFSLLVSPINCWKSMKALYNRPWRGLVLSKI----- 51     :       :
Db		4 EVIVSQYKSKSYFSHTFYLYKEI-----VNNSP-DWLHLISRLGLFLGLFAIVMHVS 54     :       :
Qy		52 ----KYKLDQLBKILELROLETISTKYNVSKQPKNUTTMETPEPYIIAGPINQSIYTL 108     :       :
Db		55 NLIKYTSDPLE-----AFKTVNRHNWSDEQREHF-----YDLRN 89     :       :
Qy		109 WFDPFYSTOLRRKPakyvysqynhtaktitfrpppcgrvpstmlcsemlnvskrndtgeqc 168     :       :
Db		90 YTSFCQTNL-----SLDCFTQILTNTVFNSWDIRDSC 120     :       :

```

Qy 169 GNFTFPMFFNVRWNTKLYVGPTKVN-----VDSQTIYFL----- 205
Db 121 KSAVLSPL-QNLPRTEIKVLSSTANKSIIASSFSILFYLLPATLTYTADPPCVLLP 179
Qy 206 -----GLTALLLYAQRNCHTSFYLVNMSRNLFRVPKYNGTKLQNTM 249
Db 180 FKILGAQLFDIKLTEESLRMSKFSNLSR--LTSFTSKFNFTSYFY-FLLYNT- 235
Qy 250 RKLKRAQPVKEQFEKAKK-----TQS-----TTTPYFSVTSAL----- 286
Db 236 ----TSCVPSNDQYFKQSPKPIVNTTSFGRVAINFDSILITTP---SSTSASLSPHPS 288
Qy 287 -NVNTNVTYSITTAARVSTSIAYRDPSSPMKIMATQLR-----DLATWVYT----- 334
Db 289 TNITPAPPVTKNSKLTHTIKVTNTPIITQTIESIKKIVKRSDFPRPMTPTDIP 348
Qy 335 TLRVQRNPFCSRRNRRAVSEFMKNTVHLINETPYIYGLDMSSLYYNMETMVEKNT- 393
Db 349 TLTLRLNATIKTEQNTENPKSPKPTNF---ENTTIRPKTLESATATTNATQKISTTF 405
Qy 394 ----ASDSNKTPTSP-----SMGQRT--FID-----PLWDYLDLSLLFLDIRNPSL 435
Db 406 TTIGIKINGNTYSSPKSNIYLSKQSQSTTKFTDAEHTTPI---LKFTWQNTVRTYMS 462
Qy 436 RSPTVNLTPPEHRRVNLSTSNL 460
Db 463 HNTVQWNTDKFOR--TTLKSSNEL 485

RESULT 7
E71611
hypothetical protein PF0580w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: E71611
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: E71611
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1224 <GAR>
A:Cross-references: UNIPROT:O96209; GB:AE001404; GB:AE001362; NID:g3845219; PIDN:AACT190
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF0580w

Query Match 4.4%; Score 108.5; DB 2; Length 1224;
Best Local Similarity 21.7%; Pred. No. 6;
Matches 78; Conservative 47; Mismatches 116; Indels 119; Gaps 19;

Qy 73 KYNVSKOPVKNLWNTE--FPQYI-LAGPQNTYSITLW-----FDP----- 112
Db 82 KYRNNNNVKNWILKDEILDYXNLHLSNLHNDINFIYSSNNINLNNKNNKPYENS 141
Qy 113 -YSTQLRKPAK-----YVYSQNHHTAKTITRPPCPGVRPSMTCLSEMLN-----VSKRN 161
Db 142 KNSNEIKKHKNVKNVYNTNHHYTKKKNKFNSTPT-EVNTNLSLNNKLNKNSLYSFRK 200
Qy 162 DTGEGCGNFTTPEMPFNVRWNTKLYVGPTKYNVDSQTIYFLGLTALLRYAQRNCHT 221
Db 201 DT-----SNFN---PSCDKNNT-----TFSKPNCLH 223
Qy 222 SFYLVNMSR-----NLFRVPKYNGTKLQNTMRKLR- KOAPVKEQFEKAKKTQSTTT 275
Db 224 E---SNPSTCTCPVNVNTIPLAIN--LLNNVNDDISPIHPLPSESSTASSTASAST 278
Qy 276 PYFSYTTSAALNVTNVTYSITTAARVSTSTIAYRDPSSPMKIMATQLRDLATWVYT 335
Db 279 ---SASTSASTSVSTSVSTSA---STTMNSPRPSDNHISNFSPLRESRATEQVNR 332

```

```

Qy 336 LRY-----RQNPFCBSPSRNRRTAVSEPMKNTVHLIRNETPYTI 372
Db 333 LYFPVNDVTSKSDPNPNNELTSNNMPKHEICEBTRN-----DNHG--IRNNSIYPL 382

RESULT 8
T16693
hypothetical protein R05F9.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16693; T27894
R:Hallsworth, K.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid R05F9.
A:Reference number: Z18559
A:Accession: T16693
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-919 <HAL>
A:Cross-references: UNIPROT:Q21750; EMBL:U41533; NID:g1109820; PIDN:AAA831
R:Hallsworth, K.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid ZK546.
A:Reference number: Z20437
A:Accession: T27894
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-575 <HA2>
A:Cross-references: EMBL:U29380; PIDN:AAA68743.1; CESP:ZK546.10
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:R05F9.12; CESP:ZK546.10
R:Introns: 31/3; 94/2; 174/3; 329/3; 524/3; 575/2; 607/2; 645/2; 739/1; 816/3
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homology

Query Match 4.4%; Score 107.5; DB 2; Length 919;
Best Local Similarity 19.2%; Pred. No. 4.8;
Matches 102; Conservative 67; Mismatches 210; Indels 153; Gaps 24;

Qy 58 LKLEILLRQLSTTSTKYNVSKQPVKNLTM-----NTEFPQYIILAGPQNTYSITV----- 107
Db 374 IEWETKAQVTAQNLYPMAKDTKIMLVGVWPDNHNHVAFPDFLDSTNNNTQNNWINEFVNYQ 433
Qy 108 -----LWPDF-----YSTQLRKPAKYVSYQNHTAKTITFRP-----PPCGRVPSMT-- 149
Db 434 SQVAFDGIWIDMNEPSNFGTNQDHP--WYFDSDDHPNDAPLFCPTNGSSSPWEMPYKTRA 491
Qy 150 -----CLSEMLNVSQR---NDTGEQCGNFTTFNP-----MFF 179
Db 492 VWRFGDANGAFLSSNTLCMLAQODGKQRFYNATGKRVAVSRSTYPSAGRYAGHWLGD 551
Qy 180 NVPRWN--TKLYVGPTKVN-----VDSQTIYFLGLTA--LLRYAQRNCHTSFY----L 225
Db 552 NTARWEDLRTSVTGAQEFNLFIPYVGSVDVCGFTTTEELCLRWQMGAFSPFRNNT 611
Qy 226 VNAMSRNLFRVPKYNGTKLQNTMRKLRKOAPVKEQFEKAKKTQSTT--TPYFSYTT 283
Db 612 IGAPQADPAWPSVAATKANLFR---YQVLPVLSLHFTASLSGATVIRPVFEYPTD 668
Qy 284 AALNVTNVTYSITTAAR-----RVSTSTIAYRDPSSPMKIMATQLRDLATWVYT 335
Db 669 AE---TFNLGYEWMGSRILVAPVIYQTTSVNAYLPTD-----RW-YSL 709
Qy 336 LRYQRNPFCE-----PSRNRVAVSEPMKNTVHLIR-----NETPYIY----- 373
Db 710 FDYRYSIMSGYATVPAPTTSRIPVVRGYSVIPROTPTSTTTTATRSNPFELIAPCPL 769
Qy 374 ----GTU-----DMGSLYYNETMFVENKTASDNKTTPTS-----PSMGFOR 411
Db 770 GMGEGTLYWDGGETIVNDFNSYDHQDFWYNSTATGCVTIITHSKSKSLSLFTLDIE 829
Qy 412 TFDPLWDYLDLSLLFLDIRNFSRSTYVNLTPPEH---RRAVNLSTSNL 460

```







Db 270 SPSTSSVATSPVPSALSSISSTPFMKPSIPTIPTIPSAYSASVSQPPPLTHSYVHPG 329  
Qy 445 PPEHRAVNLSTNSLW 461  
Db 330 PQSHKYSLSGPPASLY 346

RESULT 14  
S38181  
floculation protein FLO1 homolog YKR102w - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
C:Accession: S38181  
R:Gallion, L.; DuJon, B.  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38175  
A:Accession: S38181  
A:Molecule type: DNA  
A:Residues: 1-1169 <GAI>  
A:Cross-references: UNIPROT:P36170; EMBL:Z28327; NID:G486608; PIDN:CAA82182.1; PID:G4866  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:FLO10  
A:Cross-references: SGD:S0001810; MIPS:YKR102w  
A:Map position: 11R

Query Match 4.1%; Score 101.5; DB 2; Length 1169;  
Best Local Similarity 25.6%; Pred. No. 19;  
Matches 52; Conservative 31; Mismatches 91; Indels 29; Gaps 7;

Qy 269 KTGTTTTPFYTTSAALNVNTVYSITTA--ARRVSTSTIARPDSSFMKS-----I 320  
Db 471 ESTSTSTPYVTSSSSSEVCTECTESTSYVTYPVSSSTAAANYTSSFSSEVCTEC 530  
Qy 321 MATQLRLDWTWYTLRYQNPFCPSNRRTAVSEFMKNTHLIRNETPYTIYGLD--- 377  
Db 531 TETESTSTSTPYVTSSSSSEVCTECTESTSYVTYPVSSSTAAANYTSSFSSEVCTEC 579  
Qy 378 MSSLYNNTMFE-NKTASDSNKTTPTSPSMGFQRTFDPLMDYLDLSLLFLDEIRNFSLR 436  
Db 580 TSSPSSSEVCTECTESTSTSTPYATSTGTATSTASTNTMTSLVQDITTVSFSLS 639  
Qy 437 SPTVNLTPPEHRAVNLSTNS 459  
Db 640 S-----TVSEHTNAPTSSVESN 656

RESULT 15  
A44140  
cellulose-binding protein A precursor - Clostridium cellulovorans  
C:Species: Clostridium cellulovorans  
C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 09-Jul-2004  
C:Accession: A44140  
R:Shoseyov, O.; Takagi, M.; Goldstein, M.A.; Doi, R.H.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3483-3487, 1992  
A:Title: Primary sequence analysis of Clostridium cellulovorans cellulose binding protein  
A:Reference number: A44140; MUID:92228810; PMID:1565642  
A:Accession: A44140  
A:Status: Preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-1848 <SHO>  
A:Cross-references: UNIPROT:P38058; GB:I73817; NID:g144746; PIDN:AAA23218.1; PID:g144749  
C:Genetics:  
A:Gene: cbpA

Query Match 4.1%; Score 101.5; DB 2; Length 1848;  
Best Local Similarity 19.0%; Pred. No. 35;  
Matches 101; Conservative 77; Mismatches 173; Indels 181; Gaps 25;

Qy 46 VLSKIGYKLDQLKLEILRLQLETTISTKYNVSKOPVKNLWTNTTTPQYI----- 94  
Db 291 VTATIGKQVQN-----AGETAVPVNLTQVPAAGLA-TIELPLTFDSASLEVSIT 340

Qy 95 ---ILAGPIQNYSS-----ITYLWFD---FYSTQLRKPAKYVYVSYQYNHTAKTIT----- 136  
Db 341 AGDIVLNPSSVNFSTVSGSTIKLLFDDTLGSQLITKDG--VFATITPKAKAITTTAKV 398  
Qy 137 -----FRPPPCGRVPMTCLSEMLNVSKRN-DTGEQGCNFTTFNPMFFN 180  
Db 399 TSVKLAGTPPVGDAQLQEKPCAVNPQVTVINPIDRMQISVGTATVKAGEIAAIVPTLTS 458  
Qy 181 VPRNWTLYVGPPTKVNVDSTQIYFPLGTA--LLLYAQRNCTHSP-YLVNA-MSRNLF-- 234  
Db 459 VP--STGIATAEAQVSPDATLLEVASVTAGDIVL-----NPTVNFSTVTVNGNVIKLLFLD 511  
Qy 235 -----RVPKYNGTKL--KNTMRKLRKQA----- 257  
Db 512 DTGLSQLISKDGVFVTFNFKAKAVTSTVTTPTVTVSGTPVPADGTLAEVQSKTAAGSVTIN 571  
Qy 258 -----PVKEQPEKAKAKTQSTTTPYFSTYTTSAALNVNTVYTSITTAARRVSTS 306  
Db 572 IGDFILEPTISPTATFDKKAPADVATMTLNGYTFNGITGLTTS-DYSISGNVVKISQA 630  
Qy 307 TIAYRPPDSSFMSIMATQLRDLATWYVYTLRYQRNPFCEPSNRRTAVSEFMKNTHLIRN 366  
Db 631 YLAKQP-----VGDLE-----TLTFNFS-----NGNKTATAKL-----VWSIK 662  
Qy 367 ETPYTIYGLDMSSLYNNTMFEVNTKASDSNKTTPTSPSMGFQRTFD----- 415  
Db 663 DAPKTVTATVGTATVNAGETVAVP-VTLSNVSGISTAEALQLSPDATLLEVVSVITAGDIVL 721  
Qy 416 -PLWDYLD-----SLLFLDEIR-----NFSLRSPYVNLTP 445  
Db 722 NPSVNFSSVNGSTIKLLFDDTLGSQLISKDGVFATINFKAKSVTSVTTTP 773

Search completed: December 7, 2004, 15:21:27  
Job time : 47 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:03:02 ; Search time 154 Seconds  
(without alignments)  
1083.176 Million cell updates/sec

Title: US-09-942-146A-1

Perfect score: 2451

Sequence: 1 MGRKEMVRDVPKMFVLISI.....PEHRRVNLSTNSLMMWLQ 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_23Sep04.\*

- 1: geneseqp1980s.\*
- 2: geneseqp1990s.\*
- 3: geneseqp2000s.\*
- 4: geneseqp2001s.\*
- 5: geneseqp2002s.\*
- 6: geneseqp2003as.\*
- 7: geneseqp2003bs.\*
- 8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2451	100.0	465	8	AD015374
2	2419	98.7	465	7	AD015374
3	111.5	4.5	2590	8	AD028932
4	110.5	4.5	1304	6	ABP72189
5	110	4.5	2172	4	AB067958
6	109	4.4	1091	8	ADN05348
7	108.5	4.4	1224	3	AA18258
8	108	4.4	1327	5	AB097286
9	107.5	4.3	1167	2	AAW16326
10	106.5	4.3	1167	2	AAW10653
11	106.5	4.3	1167	2	AAW31504
12	105	4.3	452	7	ADP07589
13	103.5	4.2	1062	2	AAW22721
14	101.5	4.1	1042	6	ABR57376
15	101	4.1	2481	8	AD028930
16	101	4.1	2570	4	ABG06375
17	100.5	4.1	1224	6	ABR57375
18	100.5	4.1	4315	5	ABP43908
19	100.5	4.1	5703	8	ADL23265
20	99.5	4.1	501	2	AAW82571
21	99.5	4.1	978	2	AAV17509
22	99.5	4.1	1349	5	AB047323
23	99	4.0	903	4	AB064101
24	98.5	4.0	622	6	ABP77408
25	98.5	4.0	622	6	ABU37421

26	98.5	4.0	4134	2	AAV31946
27	98	4.0	234	3	AAG06725
28	98	4.0	234	3	AAG40390
29	98	4.0	313	3	AAG06724
30	98	4.0	313	3	AAG40389
31	98	4.0	393	3	AAG40388
32	98	4.0	393	3	AAG06723
33	98	4.0	393	5	AB091164
34	97.5	4.0	638	7	ADL90059
35	97.5	4.0	992	8	ADP04643
36	97.5	4.0	1365	6	ABP72194
37	97.5	4.0	2404	4	AB069209
38	97	4.0	621	6	ABU37812
39	97	4.0	1007	4	AB065680
40	97	4.0	2000	4	AA061853
41	96.5	3.9	347	7	ADJ68384
42	96.5	3.9	698	8	ADP99077
43	96	3.9	706	2	AA075647
44	96	3.9	859	5	ABP27958
45	96	3.9	1296	5	ABG66756

#### ALIGNMENTS

##### RESULT 1

ID	AD015374	standard; protein; 465 AA.
XX	AD015374;	
AC	AD015374;	
XX	01-JUL-2004	(first entry)
DT	01-JUL-2004	(first entry)
XX	CMV glycoprotein O (gO) polypeptide.	
DE	CMV glycoprotein O (gO) polypeptide.	
XX	CMV, glycoprotein O; gO; glycoprotein H ; glycoprotein L; glycoprotein B;	
KW	CMV infection; tumour; cancer; virucide.	
XX	Cytomegalovirus.	
OS	Cytomegalovirus.	
XX	US2004013682-A1.	
PN	US2004013682-A1.	
XX	22-JAN-2004.	
PD	22-JAN-2004.	
XX	26-JUN-2002; 2002US-00942146.	
PF	26-JUN-2002; 2002US-00942146.	
XX	29-JUL-1999; 99US-0146180P.	
PR	29-JUL-1999; 99US-0146180P.	
XX	28-JUL-2000; 2000US-00627986.	
PA	(COMP/) COMPTON T.	
XX	(HUBE/) HUBER M T.	
PI	Compton T, Huber WT;	
XX	WPI; 2004-121532/12.	
DR	Designing new anti-CMV drug by analyzing binding of glycoprotein O to	
XX	glycoprotein O receptor, designing candidate drug that would	
PT	competitively interfere with glycoprotein O binding to glycoprotein O	
XX	receptor.	
PS	Disclosure; Fig 3; 7pp; English.	
XX	The invention relates to a method of designing a new anti-cytomegalovirus	
CC	(CMV) drug, involving analysing the binding of glycoprotein O to a	
XX	glycoprotein O receptor, designing a candidate drug that would	
CC	competitively interfere with glycoprotein O binding to the glycoprotein O	
XX	receptor and showing that the candidate drug competitively inhibits	
CC	glycoprotein O binding to the glycoprotein O receptor. The invention also	
XX	relates to a method of screening involving determining whether a	
CC	candidate drug interferes with a glycoprotein O-containing complex	
XX	binding to a cell surface. The evaluation involves analysing the	
CC	inhibition of major early protein of the human cytomegalovirus (HCMV).	

```
CC The invention also relates to a vaccine useful for diminishing CMV
CC infection and comprising at least a fragment of the glycoprotein O
CC polypeptide in combination with a carrier. The vaccine comprises at least
CC a fragment of a glycoprotein chosen from CMV glycoprotein H, CMV
CC glycoprotein L or CMV glycoprotein B, and is useful for vaccinating a
CC patient against CMV infection. The drug and the vaccine are useful for
CC treating CMV infections in tumours. This sequence represents the CMV
CC glycoprotein O (gO) polypeptide of the invention.
XX
SQ Sequence 465 AA;

Query Match 100.0%; Score 2451; DB 8; Length 465;
Best Local Similarity 100.0%; Pred. No. 3.3e-218;
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGRKEMVRDVPKMFVLISISFLVSVFVCKVMSKALYNRPWRLGLVLSKIGYKLDQLKL 60
DB 1 MGRKEMVRDVPKMFVLISISFLVSVFVCKVMSKALYNRPWRLGLVLSKIGYKLDQLKL 60

QY 61 EILRQLETTISTKYNVSKQPVKNTWNTPEPQYVILAGPIQNYISITVLMWDFYSTQLRKP 120
DB 61 EILRQLETTISTKYNVSKQPVKNTWNTPEPQYVILAGPIQNYISITVLMWDFYSTQLRKP 120

QY 121 AKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
DB 121 AKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180

QY 181 VPRWNTKLVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKYI 240
DB 181 VPRWNTKLVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKYI 240

QY 241 NGTKLKNTRKLRKQAPVKQPEKAKKTOSTTTPYFSYTTSAALNVTNTVYSITTA 300
DB 241 NGTKLKNTRKLRKQAPVKQPEKAKKTOSTTTPYFSYTTSAALNVTNTVYSITTA 300

QY 301 RVSTSTIAYRPDSSFMKSIMATQLRDLATWVYTLRYQRNPFCPSRNRRTAVSEFMKNT 360
DB 301 RVSTSTIAYRPDSSFMKSIMATQLRDLATWVYTLRYQRNPFCPSRNRRTAVSEFMKNT 360

QY 361 HVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFORTFIDPLWDY 420
DB 361 HVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFORTFIDPLWDY 420

QY 421 LDSLFLDLEIRNFSLSRSPYVNLTPPEHRAVNLTSTNSLWMLQ 465
DB 421 LDSLFLDLEIRNFSLSRSPYVNLTPPEHRAVNLTSTNSLWMLQ 465

RESULT 2
ADB81330
ID ADB81330 standard; protein; 465 AA.
XX
AC ADB81330;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cytomegalovirus glycoprotein O polypeptide.
XX
KW glycoprotein O; gO; anti-cytomegalovirus; CMV; AIDS;
XX bone marrow transplant; glycoprotein H; glycoprotein L; gene therapy.
XX
OS Human herpesvirus 5.
XX
FN US6569616-B1.
XX
PD 27-MAY-2003.
XX
PF 28-JUL-2000; 2000US-00627986.
XX
PR 29-JUL-1999; 99US-0146180P.
XX
PA (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
```

```
PI Compton T, Huber MT;
DR WPI; 2003-605462/57.
XX
PT Screening candidate anti-cytomegalovirus (CMV) drugs for the ability to
PT block CMV entry into a host cell comprises determining whether a
PT candidate drug interferes with glycoprotein O binding to glycoprotein O
PT receptor.
XX
PS Disclosure; Fig 3; 8pp; English.
XX
CC This invention relates to a novel method of designing a new anti-
CC cytomegalovirus (CMV) drug that has the ability to block CMV entry into a
CC host cell. CMV is a member of the Herpesviridae family and is an
CC opportunistic pathogen responsible for serious clinical disorders in
CC immunosuppressed patients, such as persons with AIDS or those with recent
CC organ or bone marrow transplants. The present invention describes a new
CC drug target that comprises a CMV viral component designated glycoprotein
CC O (gO), whereby the candidate drug works by interfering with gO binding
CC to its receptor, where glycoprotein L. Accordingly, using gene therapy or
CC glycoprotein H and glycoprotein L. Furthermore, it is possible to block
CC developing antibodies to gO that act as vaccines, it is possible to block
CC virus infection by impeding the viral entry pathway. Furthermore, the
CC method is useful in identifying alternative drug targets and immunogens
CC that elicit protective activity against CMV infection. This polypeptide
CC is the human cytomegalovirus glycoprotein O amino acid sequence of the
CC invention.
XX
SQ Sequence 465 AA;

Query Match 98.7%; Score 2419; DB 7; Length 465;
Best Local Similarity 99.4%; Pred. No. 3e-215;
Matches 463; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 1 MGRKEMVRDVPKMFVLISISFLVSVFVCKVMSKALYNRPWRLGLVLSKIGYKLDQLKL 60
DB 1 MGRKEMVRDVPKMFVLISISFLVSVFVCKVMSKALYNRPWRLGLVLSKIGYKLDQLKL 60

QY 61 EILRQLETTISTKYNVSKQPVKNTWNTPEPQYVILAGPIQNYISITVLMWDFYSTQLRKP 120
DB 61 EILRQLETTISTKYNVSKQPVKNTWNTPEPQYVILAGPIQNYISITVLMWDFYSTQLRKP 120

QY 121 AKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 AKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180

QY 180 NVRWNTKLVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 239
DB 181 NVRWNTKLVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 240

QY 240 INGTLKNTWRLKLRKQAPVKQPEKAKKTOSTTTPYFSYTTSAALNVTNTVYSITTA 299
DB 241 INGTLKNTWRLKLRKQAPVKQPEKAKKTOSTTTPYFSYTTSAALNVTNTVYSITTA 300

QY 300 ARVYSTIAYRPDSSFMKSIMATQLRDLATWVYTLRYQRNPFCPSRNRRTAVSEFMKN 359
DB 301 ARVYSTIAYRPDSSFMKSIMATQLRDLATWVYTLRYQRNPFCPSRNRRTAVSEFMKN 359

QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFORTFIDPLWD 419
DB 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFORTFIDPLWD 419

QY 420 YLDSLLFLDLEIRNFSLSRSPYVNLTPPEHRAVNLTSTNSLWMLQ 465
DB 420 YLDSLLFLDLEIRNFSLSRSPYVNLTPPEHRAVNLTSTNSLWMLQ 465

RESULT 3
ADO28932
ID ADO28932 standard; protein; 2590 AA.
XX
AC ADO28932;
XX
```



XX The present sequence is the protein sequence of novel secreted protein  
CC PMSR from the mouse malaria parasite, Plasmodium berghei. The invention  
CC provides novel modular secreted proteins from Plasmodium falciparum  
CC (PMSR), P. berghei (PMSR) and Plasmodium yoelii (PMSR) that are post-  
CC translationally modified in sporozoites. These include Limulus clotting  
CC factor, lipid binding, scavenger receptor cysteine-rich and pentraxin  
CC domains. Homologues were identified in Plasmodium knowlesi (PKSR),  
CC Plasmodium chabaudi (PCSR), Cryptosporidium parvum (CpSR), Toxoplasma  
CC gondii (TgSR) and Theileria parva (TpSR). PMSR, PMSR and PMSR are from a  
CC unique and previously unrecognized family containing domains associated  
CC with binding and modulating host proteins involved in immunity. They are  
CC the first proteins from Plasmodium which are secreted and which show a  
CC gene disruption phenotype in both vertebrate and mosquito, making them  
CC highly useful for the production of malaria vaccines. Transgenic PMSR  
CC null mutants have attenuated growth in immunocompetent mice. They fail to  
CC produce sporozoites in Anopheles stephensi mosquitoes and consequently  
CC are dead-end parasites in the midgut of the mosquito. The novel  
CC Plasmodium proteins, and the nucleic acids encoding them, are used in  
CC claimed methods for the detection/diagnosis of malaria, and are also used  
CC in vaccine compositions. Agents capable of antagonising, inhibiting or  
CC interfering with the function or expression of the secreted protein are  
CC used in the treatment and/or prophylaxis of malaria

XX Sequence 1304 AA;  
Query Match 4.5%; Score 110.5; DB 6; Length 1304;  
Best Local Similarity 20.2%; Pred. No. 1.6;  
Matches 79; Conservative 42; Mismatches 129; Indels 141; Gaps 15;  
QY 122 KYVSYQNHAKTITFRPPCGRPVSMTCLEML-----NVSKND-TGQCGGNF--- 171  
DB 873 KYIRTLSTFT-IHFIPNSGKMKWTILSHSLCEGISIDBENELVIBQNCNPHLVK 931  
QY 172 TTENPMF-----ENVPRWNTKLYVGPVKVNDVSO-----TIY 203  
DB 932 TFGPKEHPCHELVLYNKPNSISLYNQKINLEKMKFDTLNGDLTIGRSNKQATDY 991  
QY 204 FLG-----ITALLRYAQRNCTHSPVLYNMGSRNLFPRVPKYINGTKLNTWRKL 252  
DB 992 FIGDINFVKYKILTEQIKESYDLSVNLNDGMSGN-----RDINTKTKQKTKN 1046  
QY 253 KKKQ-----APVKEQEKAKKTQSTTPVSYTTSAAALNVTTN-----VTYS 295  
DB 1047 NRKTDGRDCITPCSKTNVK-KNQVINTPEYLDSCNLLSERFNGKIGQFLVSCLED 1105  
QY 296 ITTAARRVSTSIAYRPPSSFMKSIMATQLRDLATVWYVTTLYRONPFCPSRNRVAVSE 355  
DB 1106 CTNSKYIVKGSNNYTFDTSICKAVM-----HSGIMH 1137  
QY 356 FMKNTHLVRNETPYTIYGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFD 415  
DB 1138 KTRNTH-----KNEHEDNNKNTSNS-----FLIKVE 1166  
QY 416 PLWDYLDLSLFL-----DBIRNFSLS 437  
DB 1167 GLTEYKSRGHGIVSKPEKQSLRSFSLFS 1197

RESULT 5  
ABB67958  
ID ABB67958 standard; protein; 2172 AA.  
XX  
AC ABB67958;  
XX  
DT  
XX  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 30666.  
DE Drosophila melanogaster polypeptide; cell signalling; insecticide;  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX Drosophila melanogaster.  
OS

XX WO200171042-A2.  
PN  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI: 2001-656860/75.  
XX N-PSDB; ABL12061.  
DR  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Disclosure; SEQ ID NO 30666; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at fcp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 2172 AA;  
Query Match 4.5%; Score 110; DB 4; Length 2172;  
Best Local Similarity 19.4%; Pred. No. 3.6;  
Matches 87; Conservative 62; Mismatches 174; Indels 126; Gaps 17;  
QY 67 ETTISTKYNVSK-----QPVKNTMTNTEFFQYIYLAGPIQNYISITYLWDFYSTQ-----LR 118  
DB 1450 KTSSTSYTTESTYIQTSTSYLNAEVTSH---PTTSTYEVNTIITDESKIDMFNLI 1506  
QY 119 KPARYVSYQNHAKTITFRPPCGRPVSMTCLEMLNVSKRNDTGEQCGNETTEN 175  
DB 1507 KGIEDLIQYNGKHKEFTTSLTGKNSREPTWTYSSAGGITSH-----SFTTIS 1557  
QY 176 PMFPNVPWNKLYVGP-----KVNVDSTQIYFLGLTALLRYAQRNCTHSPVLYNA 228  
DB 1558 PKTVTSSSHSTTESSESHPTTSTNYEVTVFDQD-----SHKFKVMKN 1599  
QY 229 MSRNL-----FRVPKYINGTKLNTWRKLKRQAPVKEQEKAKK----- 269  
DB 1600 LIKELKGLIEHFSDTKNITSTYTGDSKSFQPTATTPDSDRSSQPIPFPESTTTTN 1659  
QY 270 TQSTTTPYFS-YTTSAAALNVTTNTVYISITTAARVSTSIAYR-----PDSSFMKSIMAT 323  
DB 1660 LQSTITNHLSTESTSQNLTLTIT---GTTTNTLTSTHDFKQIQNDKSKLKHU--- 1713  
QY 324 QLRLATWVVT-----TLRYRONPFCPSRNRVAVSEFMKNTHLVRNETPY 370  
DB 1714 -IKTLDVLYSLNDKNNIKSTSLTSTFNSSINITNSTVASTSTSTTESMAIHSTPY 1772  
QY 371 TIYGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFDPLWDYLDLSLLFDEI 430  
DB 1773 --LSTREITVFQND---VENINPIISNNSTVTT----- 1802  
QY 431 RNFSLRSPYVNLTPPEHRRVAVNLSTNS 459  
DB 1803 ---NIHSSYTTTSTTDNTSYPNLSTTES 1828



RESULT 6

ADN05348  
ID ADN05348 standard; protein; 1091 AA.  
XX  
XX AC ADN05348;  
XX  
XX DT 01-JUL-2004 (first entry)  
XX  
XX DE Antipsoriatic protein sequence #846.  
XX  
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO2004028479-A2.  
XX  
XX PD 08-APR-2004.  
XX  
XX PF 25-SEP-2003; 2003WO-US030907.  
XX  
XX PR 25-SEP-2002; 2002US-0414006P.  
XX  
XX PA (GETH ) GENENTECH INC.  
XX  
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
XX  
XX PI Wu TD;  
XX  
XX DR WPI; 2004-305105/28.  
XX  
XX DR N-PSDB; ADN05347.  
XX  
XX PT New PRO nucleic acid or polypeptide, useful for preparing a  
XX  
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a  
XX  
XX PT mammal.  
XX  
XX PS Claim 9; SEQ ID NO 1742; 3069pp; English.  
XX  
XX CC The invention relates to novel polynucleotide and polypeptides for  
XX  
XX CC treating psoriasis or a sequence having at least 80% identity to the  
XX  
XX CC above sequences. The nucleic acid is useful for preparing a composition  
XX  
XX CC for diagnosing or treating psoriasis in a mammal. This sequence  
XX  
XX CC corresponds to one of the polypeptides of the invention.  
XX  
XX SQ Sequence 1091 AA;  
Query Match 4.4%; Score 109; DB 8; Length 1091;  
Best Local Similarity 20.7%; Pred. No. 1.7;  
Matches 84; Conservative 60; Mismatches 167; Indels 94; Gaps 20;  
Qy 55 LDQLKLEILRQLETTISTKYNVSKQPVKNL-TMNTFPPQYILAGPIONYSITYLWDFD- 112  
Db 540 LDPLDAELENDIKVEIRKMKIDGESGKTFRTLVKSQDERVIDKG-----NRTVTWTPVN 594  
Qy 113 ---YSTQLRKPAKYVSYNHTAK---TITFRPPCGRPVMTCLSEMLNVSKNDTGEQ 166  
Db 595 GTDYSALVLP---TYSFYIKAKLEETITQARSKGKMD-----SETLKPDNFEESG-- 645  
Qy 167 GCGNFTTPNPFNVPRNNTKLYGPTKVNDSDOTIYPLGLTALLRVAQRNCTHSFVLV 226  
Db 646 ----YTFIAP-----RDYCNLDKIS-DNNTFELNPNFIDRKTNPNSCNADLI 690  
Qy 227 -----NAGRNLRFPKPYNGTKLKNM-----RKLKRKQAPVKEQ 262  
Db 691 NRVLDDAGFTNELQVNVYSKQNIKGKVPVTDWMDYQSLSQGRWKLARKPETVEDS 750  
Qy 263 PEKAKKTQS---TTPPYSTTSAL--NVTNVTYSITTAARVUSTIAYRDP-SSFM 317  
Db 751 FYKSLDNDNVTFAPYFNKSGPGAYESGIMVSKAVEIYIQGLKPAVVGKIDVNSWI 810  
Qy 318 KSIMATOLDLATWYTTTLRYQRNPFCEPSNRRTAVSEFM-----KNTHVLIRNETPYT 371  
Db 811 ENFTKTSIRDPCA-----GPVCDCKN-----SDVMDCVLDDGGFLMANHDDYT 856  
Qy 372 ----YGTLDWSSLY--NETMFVENKTASDSNKTTP-TSPSMG 408

Db 857 NQIGRFFGEIDPSLRHLVNISVYAFNKSXDYQSVCEPGAAPKQG 901

RESULT 7

AAB18258  
ID AAB18258 standard; protein; 1224 AA.  
XX  
XX AC AAB18258;  
XX  
XX DT 07-NOV-2000 (first entry)  
XX  
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:115.  
XX  
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
XX  
XX KW antimalarial; malaria; protozoacide; infection; insecticide.  
XX  
XX OS Plasmodium falciparum.  
XX  
XX PN WO200025728-A2.  
XX  
XX PD 11-MAY-2000.  
XX  
XX PF 05-NOV-1999; 99WO-US026796.  
XX  
XX PR 05-NOV-1998; 98US-0107131P.  
XX  
XX PA (HOFF/) HOFFMAN S.  
XX  
XX PA (CARU/) CARUCCI D.  
XX  
XX PA (GARD/) GARDNER M.  
XX  
XX PA (VENT/) VENTER J C.  
XX  
XX PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX  
XX DR WPI; 2000-365347/31.  
XX  
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,  
XX  
XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
XX  
XX PT diagnosis of P.falciparum infection.  
XX  
XX PS Disclosure; Page 277-280; 577pp; English.  
XX  
XX CC The present invention describes proteins and their fragments (I) encoded  
XX  
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
XX  
XX CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
XX  
XX CC vaccines against P. falciparum infection comprising (I) or (II). (I) and  
XX  
XX CC (II) are useful for the development of vaccines against P. falciparum  
XX  
XX CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
XX  
XX CC immunogens comprising the sequences of (I), are useful in the detection  
XX  
XX CC of infection with P. falciparum. Furthermore, (I) (especially when they  
XX  
XX CC are fusions or secreted or membrane proteins) can aid the identification  
XX  
XX CC of drugs to treat or prevent P. falciparum infection, or they can be used  
XX  
XX CC to identify drug resistance in P. falciparum. Sequencing of the  
XX  
XX CC Plasmodium chromosome 2 and the subsequent identification of proteins  
XX  
XX CC encoded by it will help to expand our understanding of parasite biology,  
XX  
XX CC a process hampered by the complexity of the parasitic lifecycle, and  
XX  
XX CC provide new targets for vaccine and drug development. Parasite resistance  
XX  
XX CC to drugs and mosquito resistance to insecticides have led to a resurgence  
XX  
XX CC of malaria in many parts of the world, and there is a pressing need for  
XX  
XX CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352  
XX  
XX CC represent nucleotide and protein sequences given in the present  
XX  
XX CC invention, but which are not specifically mentioned within the  
XX  
XX CC specification  
XX  
XX SQ Sequence 1224 AA;  
Query Match 4.4%; Score 108.5; DB 3; Length 1224;  
Best Local Similarity 21.7%; Pred. No. 2.2;  
Matches 78; Conservative 47; Mismatches 116; Indels 119; Gaps 19;  
Qy 73 KYNVSKQPVKNLTWNTB--FPQYI-LAGPIONYSITYLW-----PDF----- 112  
Db 82 KYRKNNNVKNMILKDELDLYDYNHLSNHLNHDINFIYSSNNIFNLCKNNKPKFPNS 141

QY 113 -YSTQLRPAK-----YYSQYHNTAKTITPRPPCGRVSMTCLEMLN-----VSKRN 161  
 Db 142 KNSNEIKDKHKVNVYTNHHTKKNKPNYSNPT-EVNVNSLNLKNSLYISFRK 200  
 QY 162 DTGQCGCGNFTFNPMPFNVRWNTKLYVGTQVNVDSQTIYFLGLTALLRVAQRNCTH 221  
 Db 201 DT-----SNFN-----PSCDKNT-----TFSPKNCNH 223  
 QY 222 SFYLVNAMS-----NLFRVPKYNGTKLKNMRLKR-KQAPVKEQFEKAKKTQSTT 275  
 Db 224 E---SNPSSTCYPNVNTIPLAIN--LLNVNDDISPIHPLPSESSTSASTSAST 278  
 QY 276 PYFSYTTSAALNVTNTYTSITTAARRVSTSTIAVRDSSPMKSMATQLRLDLATVYTT 335  
 Db 279 ---SASTSASTSVSTSVSTSA---SITMNSPRPSDNIHSFPLSRESRATEQVNR 332  
 QY 336 LRY-----RONPCEPSRNRNAVSEFMKNTHVLRNTPYTI 372  
 Db 333 LYFPVNDVTSKSDPNPNELTSMNPKHEPICEETRN-----DNHG--IRNNSIYPL 382

RESULT 8  
 ABB97286  
 ID ABB97286 standard; protein; 1327 AA.  
 XX AC ABB97286;  
 XX DT 28-JUN-2002 (first entry)  
 XX DE Novel human protein SEQ ID NO: 554.  
 XX KW Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;  
 KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
 KW neuroprotective; antiparkinsonian; protein therapy; EST;  
 KW expressed sequence tag.  
 XX OS Homo sapiens.  
 XX PN WO200222660-A2.  
 XX PD 21-MAR-2002.  
 XX PF 10-SEP-2001; 2001WO-US026015.  
 XX PR 11-SEP-2000; 2000US-00659671.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT,  
 XX WPI; 2002-292408/33.  
 DR N-PSDB; ABN32472.  
 XX An isolated polynucleotide for treating diseases associated with its  
 PT encoded polypeptide such as cancer and multiple sclerosis.  
 XX Example 2; SEQ ID NO 554; 509pp; English.  
 XX The present invention provides the protein and coding sequences of 444  
 CC novel human proteins. These were isolated from expressed sequences tags  
 CC (ESTs). They can be used to stimulate cell growth, to regulate  
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke  
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a protein of the invention  
 XX Sequence 1327 AA;

Query Match 4.4%; Score 108; DB 5; Length 1327;  
 Best Local Similarity 22.0%; Pred. No. 2.8;  
 Matches 108; Conservative 59; Mismatches 221; Indels 102; Gaps 23;  
 QY 20 ISPLVSNFKVMSALYNRPWRGLVLSKIGYKLDQLKLEILRQLETTISTKY---NV 76  
 Db 266 IQFLMT---NEETVDKA---PPHSKIGLEKKRKRMDVSK--ITRYTEOCFSDSNCPNK 317  
 QY 77 SK-QPVKNLTMTNTEFPQYVILAGPIQNYSTIYLWDFYFSTQLRKPQYVYVQVNHAKTI 135  
 Db 318 SKMQEVDLEQNEE-----LOAVDSQYALSQVPESTDEDLSDVAFQHLIYN----- 366  
 QY 136 TFRPPPCGR-----VPSMTCLEMLNVSXRNNDTGEQCGNFTTNPMPFNVRWNTKLYVG 191  
 Db 367 ---PDKGEBESSPVHTSTFLSNTLK-KKCEESDSESPATFSTEEPSTFPCCTKCNVNF--- 419  
 QY 192 PTKVNVDSQTIYFLGLTALLR-----YAQRNCTHSFYLNVNAMSRLFRVPKYINGTKL 245  
 Db 420 REKKHLHRHMMYHLGNSHFRHLNVRPYACRECGRTFRDRNLSLLKHM--IIHOERRQKL 477  
 QY 246 KNTMRKLLK-----RKQAP-----VKEQFEKAKKTQSTTTPY-----FSYTT 283  
 Db 478 MEIRELKELODEGRSARLQPCQVGTNCPKTFVQAKTHEKDKRYCCCEECNFMVATE 537  
 QY 284 AALNVTNTYTSITTAARRVSTSTIAYR--PDSSFMKSMATQLRLDLATVYTTLYRQN 341  
 Db 538 NELECHRGIAHGAVVKCPMV-TSDIAQRKTQKTFMKDSVVGSKKSATYI-----CKMC 591  
 QY 342 PCEPSRNRNAVSEFMKNTHVLRNTPYTIYGTLDMSLYYNETMFVENKTSADSNKTT 401  
 Db 592 PF-----TTSKSVLKHTTEYLSHSSCVDSFGSPLGLDK---RKNDILEEPVDSSTKTL 643  
 QY 402 PTSPSMGFORFTDPLDWDYLDLSLLFLDEIRNF--SLRSPYVNLTPPEHR----- 449  
 Db 644 TKQOSTTFPK-----NSALKQDVKRTFSTSSQSSFSKHKHPHRIQARKASIAQ 693  
 QY 450 RAVNLTSTNS 459  
 Db 694 SGVNMCMQNS 703

RESULT 9  
 AAW16326  
 ID AAW16326 standard; protein; 1168 AA.  
 XX AC AAW16326;  
 XX DT 17-AUG-1997 (first entry)  
 XX DE Nematocidal toxin 167P.  
 XX KW Nematocide; pesticide; biological control; corn rootworm; Diabrotica;  
 KW crystal protein; CryV; endotoxin.  
 XX OS Bacillus thuringiensis.  
 XX Key Location/Qualifiers  
 XX FT Misc-difference 1168  
 XX FT /note= "the methionine residue at C-terminal position  
 FT 1168 is additional to the sequence deduced from the 167P  
 gene"  
 XX WO9712980-A1.  
 XX 10-APR-1997.  
 XX 01-OCT-1996; 96WO-US015730.  
 XX 06-OCT-1995; 95US-00540104.  
 XX 21-MAR-1996; 96US-00620717.  
 XX (MYCO ) MYCOGEN CORP.  
 XX PA

```

PI Feitelson JS;
XX WPI; 1997-226223/20.
XX N-PSDB; AAT62498.
XX New Bacillus thuringiensis isolates - polynucleotide sequences encoding
XX toxins useful for controlling nematode and coleopteran pests.
XX Disclosure; Page 31-35; 46pp; English.
XX The Bacillus thuringiensis 167P pesticidal toxin (AAW16326) can be
XX expressed in recombinant microorganisms, e.g. B.t., E. coli or
XX Pseudomonas, for use in the biological control of pests, or expressed in
XX transgenic plants to improve pest resistance. The amino acid sequence of
XX 167P was deduced from the 167P gene coding sequence (AAT62498)
XX SQ Sequence 1168 AA;

Query Match 4.4%; Score 107.5; DB 2; Length 1168;
Best Local Similarity 20.9%; Pred. No. 2.6;
Matches 97; Conservative 68; Mismatches 131; Indels 169; Gaps 26;

Qy 90 PQQYV-ILAGPIQNYSTIYLMFDFYSTQLRKPAKYVSYQVNHAKTITFRPPPCGRVPSM 148
Db 8 YPSTHNVLAHPIRLDS---FFDPF-VETFKDLKGAWEEF-----GKTGYM 48
Qy 149 TCLSEMLNVSKRNDTGEQCGNFTTFNPMFNPVRWNTKLVGPTKVNV-----SQ 200
Db 49 DPLKQHLQIAW---DTSQNG-----TVDYLAUTKASISLIGLIPGADA 88
Qy 201 TIYFGL-----TALLRYAQRNCTHSFY-----LVNMSRN--LFRVPKYNG- 242
Db 89 VVPFINMFVDFIPKLFGRGSGQNAQAFBELIIEKVKELVDEDFRNFNLNLLYLDGM 148
Qy 243 ----TKLKNTR-----KLKRKOAP-----VKQPEKAKKTOSTTTPY 277
Db 149 QTALSHFQNDVQIAICQEQGLMDQPTACTPTTDHLISVRESF-KDARTTIETALPH 207
Qy 278 FS-----YTTSAALNVNTNVTYSITTAARRVSTSIAYRPD 313
Db 208 FKNPMLSTNDNTPDPSNTVLLTLPMTYTTATLNLHLQGY-IQFAERWKSVMY-----D 261
Qy 314 SSFM---KSIWATQLRDLATVYTLRYRQNPFCPSRNRRTAVSEFMKNTVHLIRNET-- 368
Db 262 ESFINQTKVDLQRRIQDYSTTVSTTFE-KFKPTLNPS-NKESVKNYR-----YVRSMTLQ 315
Qy 369 -----PYTIYGLDMSLSLYNETMFV-----ENKTASDNKTTTSPSMG 408
Db 316 SLDIAATWPTLDNVNPSNVDIQLDQTRLVFSVAGPWEGNDNITSNIIDVLTPTINTGIG 375
Qy 409 FQ-----RTFIDPLWDYLDLSLLFLDEIRNFSLSRSPYVNLTPPEH 448
Db 376 FQESSDLRKFTYPRIE-LQSMQFHGQ-----YVNSKSVEH 409

RESULT 10
AAW10653
ID AAW10653 standard; protein; 1167 AA.
XX AC AAW10653;
XX DT 25-MAR-2003 (revised)
XX DT 22-APR-1997 (first entry)
XX DE Bacillus thuringiensis PS167P nematode active toxin.
XX KW Nematode; active toxin; biological control; animal feed; additive;
XX KW production; recombinant; resistant; transgenic plant
XX OS Bacillus thuringiensis.
XX PN US5589382-A.
XX
```

---

```

PD 31-DEC-1996.
XX 07-JUN-1995; 95US-00485568.
XX 27-JUL-1990; 90US-00558738.
XX 21-JUL-1992; 92US-00918345.
XX 31-DEC-1992; 92US-00999053.
XX 15-JUL-1993; 93US-00092155.
XX 30-DEC-1993; 93US-00176403.
XX 21-SEP-1994; 94US-00310197.
XX 16-DEC-1994; 94US-00357698.
XX (MYCO ) MYCOGEN CORP.
XX Fu J, Payne J, Narva KE;
XX WPI; 1997-076851/07.
XX N-PSDB; AAT61018.
XX Bacillus thuringiensis toxin genes - encoding toxins active against
XX nematodes.
XX Claim 4; Col 29-36; 20pp; English.
XX The present sequence is the Bacillus thuringiensis PS167P (NRRL B-18681)
XX nematode active toxin, useful in the biological control of nematodes,
XX e.g. as an animal feed additive. The toxin gene, which was isolated from
XX PS167P cellular DNA by PCR, can be used for the production of recombinant
XX toxin, and nematode resistant transgenic plants. (Updated on 25-MAR-2003
XX to correct PF field.)
XX SQ Sequence 1167 AA;

Query Match 4.3%; Score 106.5; DB 2; Length 1167;
Best Local Similarity 20.9%; Pred. No. 3.2;
Matches 97; Conservative 67; Mismatches 132; Indels 169; Gaps 26;

Qy 90 PQQYV-ILAGPIQNYSTIYLMFDFYSTQLRKPAKYVSYQVNHAKTITFRPPPCGRVPSM 148
Db 8 YPSTHNVLAHPIRLDS---FFDPF-VETFKDLKGAWEEF-----GKTGYM 48
Qy 149 TCLSEMLNVSKRNDTGEQCGNFTTFNPMFNPVRWNTKLVGPTKVNV-----SQ 200
Db 49 DPLKQHLQIAW---DTSQNG-----TVDYLAUTKASISLIGLIPGADA 88
Qy 201 TIYFGL-----TALLRYAQRNCTHSFY-----LVNMSRN--LFRVPKYNG- 242
Db 89 VVPFINMFVDFIPKLFGRGSGQNAQAFBELIIEKVKELVDEDFRNFNLNLLYLDGM 148
Qy 243 ----TKLKNTR-----KLKRKOAP-----VKQPEKAKKTOSTTTPY 277
Db 149 QTALSHFQNDVQIAICQEQGLMDQPTACTPTTDHLISVRESF-KDARTTIETALPH 207
Qy 278 FS-----YTTSAALNVNTNVTYSITTAARRVSTSIAYRPD 313
Db 208 FKNPMLSTNDNTPDPSNTVLLTLPMTYTTATLNLHLQGY-IQFAERWKSVMY-----D 261
Qy 314 SSFM---KSIWATQLRDLATVYTLRYRQNPFCPSRNRRTAVSEFMKNTVHLIRNET-- 368
Db 262 ESFINQTKVDLQRRIQDYSTTVSTTFE-KFKPTLNPS-NKESVKNYR-----YVRSMTLQ 315
Qy 369 -----PYTIYGLDMSLSLYNETMFV-----ENKTASDNKTTTSPSMG 408
Db 316 SLDIAATWPTLDNVNPSNVDIQLDQTRLVFSVAGPWEGNDNITSNIIDVLTPTINTGIG 375
Qy 409 FQ-----RTFIDPLWDYLDLSLLFLDEIRNFSLSRSPYVNLTPPEH 448
Db 376 FQESSDLRKFTYPRIE-LQSMQFHGQ-----YVNSKSVEH 409

RESULT 11
AAW31504
ID AAW31504 standard; protein; 1167 AA.

```

```
XX AC AAW31504;
XX DT 07-APR-1998 (first entry)
XX DE Nematoxin toxin 167P protein.
XX KW PCR primer; amplify; nematode toxic protein; Bacillus thuringiensis;
XX KW delta-endotoxin gene; nematode pest control; Panagrellus redivivus;
XX KW 167P protein.
XX OS Bacillus thuringiensis.
XX PN WO9734926-A2.
XX PD 25-SEP-1997.
XX PF 21-MAR-1997; 97WO-US004755.
XX PR 21-MAR-1996; 96US-00590554.
XX PA (MYCO ) MYCOGEN CORP.
XX PI Payne J, Narva KE, Fu J;
XX DR WPI; 1997-480163/44.
XX DR N-PSDB; AAT89185.
XX PT Bacillus thuringiensis toxin gene - useful in recombinant hosts,
XX PT particularly plants for the control of nematodes.
XX PS Claim 4; Page 35-39; 44pp; English.
XX CC This sequence represents the protein encoded by a polynucleotide of the
XX CC invention. The polynucleotide of the invention is a sequence from a
XX CC Bacillus thuringiensis (Bt) isolate selected from PS80JG1, PS158D5,
XX CC PS167P, PS169E, PS177F1, PS177G, PS204G4 and PS204G6, that encodes a
XX CC toxin active against nematodes. This sequence represents the 167P
XX CC protein, and is a delta-endotoxin protein. The polynucleotides and toxins
XX CC can be used for the control of nematode pests such as Panagrellus
XX CC redivivus
XX SQ Sequence 1167 AA;
    Query Match 4.3%; Score 106.5; DB 2; Length 1167;
    Best Local Similarity 20.9%; Pred. No. 3.2; Mismatches 132; Indels 169; Gaps 26;
    Matches 97; Conservative 67;
Qy 90 FPQY-ILAGTQNYSTYLYWDFYSTQLRKPAKYVYQYNHTAKTITFRPPGCRVPSM 148
Db 8 YPSYHNVLHPIRLDS---FFDPP-VETFDLKGAWEEP-----GKTGYM 48
Qy 149 TCSMLNVSKRNTGEGCGNFTTFNPMFVPRWNTKLVGPTKVNVD-----SQ 200
Db 49 DPLKQHLQIAW--DTSQNG-----TVDYLAUTKASISLIGLPGADA 88
Qy 201 TIYFGL-----TALLRYAQRNCTHSFY-----LYNAMESN--LFRVPKYNG- 242
Db 89 VVPFINFVDFIPFKLFGSGQNAQAQFFELLIEKVKELVDEDFRNLNLLNLDGM 148
Qy 243 ----TKLQNTMR-----KLKRRQAP-----VKEQFKKAKKTOSTTTPY 277
Db 149 QTALSHFQNDVQIAICQEQEGLMDQPTACTPTTDHLISVRESF-KDARTTLETALPH 207
Qy 278 FS-----YTTSAALNVNTVNTVSTTAARVSTSTIARPD 313
Db 208 FKNPMLSTNDNTPDFNSDTVLLTLPMTYTTGATLNLILHOGY-IQFAERWKSVMY-----D 261
Qy 314 SSEM---KSINATQRLATWYTTTLRVQNPFCPEPSRNRATVGEFMKNTHLIRNET-- 368
Db 262 ESFINQTKVDLQRRIQDSTIVTSITFE-KFPTLNPS--NKESVKNYKIR-----YVRMTLQ 315
Qy 369 -----PYTIYGTLDMSLYNETMFV-----ENKTASDSNKTPTTSPSMG 408
```

```
Db 316 SLDIAATWPTLDNVNPSNVDIQDTRLVFSVDVAGPWEGNDNITSNIIDVLTPINTGIG 375
Qy 409 FQ-----RTFIDPLWDYLDLSLLFLDEIRNFSLSRSTYVNLTPPEH 448
Db 376 FQESSDLRKFTYPRIE-LQSNQFHGQ-----YVNSKSVEH 409
RESULT 12
ID ADF07589 standard; protein; 452 AA.
XX AC ADF07589;
XX DT 12-FEB-2004 (first entry)
XX DE Bacterial polypeptide #3702.
XX KW Proteus mirabilis infection; bacterial infection; antibacterial;
XX KW immunostimulant.
XX OS Proteus mirabilis.
XX PN US6605709-B1.
XX PD 12-AUG-2003.
XX PF 05-APR-2000; 2000US-00543681.
XX PR 09-APR-1999; 99US-0128706P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL;
XX DR WPI; 2003-895291/82.
XX DR N-PSDB; ADF03417.
XX PT New Proteus mirabilis polypeptides and polynucleotides, useful as
XX PT reagents for diagnosis of bacterial disease, as components of
XX PT antibacterial vaccines, as targets for antibacterial drugs, or as
XX PT biocontrol agents for plants.
XX PS Disclosure; SEQ ID NO 7874; 870pp; English.
XX CC The invention relates to new Proteus mirabilis polypeptides and
XX CC polynucleotides. The invention also relates to antibodies against the
XX CC polypeptides, methods for producing the polypeptides, a method of
XX CC generating vaccines for immunising an individual against P. mirabilis, a
XX CC method for evaluating a compound for the ability to bind a P. mirabilis
XX CC polypeptide and a method for screening test compounds for anti-bacterial
XX CC activity. The polypeptides and polynucleotides are useful as molecular
XX CC targets for diagnosing, preventing and treating pathological conditions
XX CC resulting from bacterial infection, as reagents for diagnosis of
XX CC bacterial diseases, as components of antibacterial vaccines, as targets
XX CC for antibacterial drugs or as bio-control agents for plants. This
XX CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX SQ Sequence 452 AA;
    Query Match 4.3%; Score 105; DB 7; Length 452;
    Best Local Similarity 21.5%; Pred. No. 1.1;
    Matches 88; Conservative 57; Mismatches 143; Indels 122; Gaps 22;
Qy 14 MFVLISFLVSPINCKVMSKALYNRPWGLVLSKIGYKLDQLKLEIL-----RQLET 68
Db 38 LYIYIIFILMEIHVNKRALSTKYRNSMYPMSI-KKIGGYGLSLFLPIIFHHSANALEL 96
Qy 69 TI---STKY-NVSKQPKVNTMTNTEFPQYIYLAGPIQ-NYS-ITYLW-----EDFY 113
Db 97 SVPIGSTGYIGVETPPVGSTTKN-----YGMVSSVLQANRSYISFRWEPQFVLSLDPFY 151
Qy 114 STQLRKPAKYVYQYNHTAKTITFRPPGCRVPSMTCLSEMLNVSKRNTGEGCGNPF-- 171
```



```
CC asparagine, X = any amino acid other than proline, and B = serine or
CC threonine) has been modified so that no N-glycosylation occurs during the
CC expression in a eukaryotic cell. Also described: (1) a fused DNA
CC molecule, where a DNA encoding a signal sequence has been ligated to the
CC N-terminal end of the modified DNA molecule as described above so that it
CC may be expressed as a fusion protein; (2) a recombinant virus integrated
CC with the DNA molecule or the fused DNA molecule described above; (3)
CC producing a modified or fusion protein by using the recombinant virus
CC described above, to express a protein encoded by the modified DNA
CC molecule or the fused DNA molecule in a eukaryotic cell; and (4) a
CC vaccine comprising the recombinant virus. The DNA molecule has virucide
CC and immunostimulant activities. The DNA molecule is useful for producing
CC a vaccine for treating viral infections. The present sequence is used in
CC the exemplification of the present invention
XX
SQ Sequence 1042 AA;
Query Match 4.1%; Score 101.5; DB 6; Length 1042;
Best Local Similarity 18.1%; Pred. No. 7.9;
Matches 114; Conservative 73; Mismatches 223; Indels 221; Gaps 28;
QY 7 MVRDVKMFVLISIFLVSFINKCKVMSKALYNRPWGLVLS-KIGKYKLDQLKLE---- 61
Db 48 VLRDVDNFITAQGTII-----KLD$---FTRPLVGLDLSDDFAGYKVKQIVSDYTT 98
QY 62 -----ILROLETTISTK-YNVSKQPVKNLTWTEP-----POY----- 93
Db 99 RNRFDQRTRAYALLVDEANVHLKINTNSNRGNQNSKFIQGVNDPAHVIRPTD 159
QY 94 -----YILAGPIQNYSIITLVWFDYSTQLRPAKYVSYQVNYHTAKTI 135
Db 159 DGTKEFQFTQGTGEIWNDFILDAPILPKDLPDWNLYIQKILP-----NDVNTAV 210
QY 136 TFRPPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTF-----NPMFFN 180
Db 211 V--PWPVGRVSGTNADGDFGNGQITNDPIAQTKTITDQNPNSTFNSGAMPGANNRY 268
QY 181 -----VPRWN-----TKLYGPTKVNVDSTQTYIFGLTALL 211
Db 269 DSQLVNKHRIKTSFQDKEKFVPEWTEGSEENKQITRLATGSLPSNERYWILIDIPGTPOVT 328
QY 212 LVAQNRNCTHSFVLNAMSRLNFRVPKYING-----TKLXNTMRKLUKRQAP 258
Db 329 LKEDSVNVPRLYLNSVNSLSFIGDSIYIFGTSPLSLWYSPFTRLSDLTALNQVKTDD 388
QY 259 VKEQFEKAKKTQSTTPVFTTGAALNVTNTVYSITTAARRVYSTTI-AYRPDSSFM 317
Db 389 IEASSTDQGTQTQGTIT-----TTDTSGSTGAGTGQTQTSTQTSVNSPILNTYRSFGIDS 443
QY 318 KSIMATQLDLATW-----VYTLRYR-QNPFCEPSRNRATAVSBFMKNT----- 360
Db 444 KPT$ANKI--DETWNADPNVIEARIYAEYRLGLQNEI--PITN---AGNFIRNTIGVGFT 497
QY 361 ----HVLIR-----NETPYT-ITYGTLD-----MSLYNYNETMFEV----- 391
Db 498 STGSRVVLASVNGDQRPTGNFQPPFLYVGLYQQTGTGTWYGYTKLLQSPYDLDA 557
QY 392 -KTASDN---KTTPTSPMG-----FRTTIDPLW 418
Db 558 ARVGTETNQFRRTSLTYPMVGGYLTBEGARSPNTPIYIRAGDTPESRIFSGYSDNTY 617
QY 419 DYLDLSLLELDEIR---NFSLRSPTYVNLTPP 446
Db 618 EYIQSVLGFDCIRNLLNVGVKASSFLNSNRP 648
RESULT 15
AD028930
ID AD028930 standard; protein; 2481 AA.
XX
AC AD028930;
XX
DT 29-JUL-2004 (first entry)
```

```
XX
DE Human novel GPCR PGRI7, SEQ ID NO:29.
XX
XX G protein-coupled receptor; GPCR; drug screening; diagnosis;
XX transgenic mouse; neurological disorder; adrenal gland disorder;
XX colon disorder; intestinal disorder; cardiovascular disorder;
XX muscular disorder; blood disorder; immune disorder; bone disorder;
XX joint disorder; metabolic disorder; nutritive disorder; cancer;
XX kidney disorder; liver disorder; lung disorder; breast disorder;
XX ovary disorder; uterus disorder; prostate disorder; spleen disorder;
XX skin disorder; stomach disorder; antiparkinsonian; antimanic;
XX thymus disorder; thyroid disorder; vasotropic; antiangiinal; antiarhythmic;
XX cytostatic; antiinflammatory; respiratory; antidiarrhoeic; antidiabetic;
XX CNS; central nervous system; antibacterial; antinaeamic; antiseborrheic;
XX virucide; hepatotropic; antitumor; antithyroid; antiallergic; anorectic;
XX dermatological; antiulcer; antihypertropic; gene therapy; GPCR modulator; human;
XX immunosuppressive; nephrotropic; receptor.
XX
XX Homo sapiens.
OS
XX WO2004040000-A2.
XX
XX 13-MAY-2004.
XX
XX 09-SEP-2003; 2003WO-US028226.
XX
XX 09-SEP-2002; 2002US-0409303P.
XX
XX 09-APR-2003; 2003US-0461329P.
XX
XX (PRIM-) PRIMAL INC.
XX
XX Gaitanaris GA, Bergmann JE, Gragerov A, Hohmann J, Li F;
XX Madisen L, McIlwain KL, Pavlova MN, Vassilatis D, Zeng H;
XX
XX WPI; 2004-390329/36.
XX
XX N-PSDB; ADO28931.
XX
XX Novel mammalian G protein coupled receptors, useful for identifying
XX compounds that modulates diagnosing and treating disease condition
XX associated with GPCR dysfunction e.g. autoimmune diseases, angina
XX pectoris, Parkinson's disease.
XX
XX Claim 1; SEQ ID NO 29; 542pp; English.
XX
XX The invention relates to human and mouse G protein-coupled receptors
XX (GPCRs) and nucleic acids encoding them. The invention also relates to
XX sequences at least 90% identical to the GPCR proteins and nucleic acids
XX of the invention; methods of treating, preventing or diagnosing diseases
XX associated with GPCRs of the invention; methods of screening for
XX compounds useful in the treatment of GPCR-related diseases; a transgenic
XX mouse comprising a GPCR gene of the invention; a mouse comprising a
XX mutation in a GPCR transgene or in an endogenous GPCR gene; cells of which has
XX from the transgenic mice; kits comprising several mice, each of which has
XX a mutation in a different GPCR gene of the invention; and kits comprising
XX probes which hybridise to GPCR polynucleotides of the invention. The
XX invention further discloses variants of the GPCR polypeptides and vectors
XX comprising a GPCR nucleic acid. The GPCR nucleic acids and proteins may
XX be used in the diagnosis, treatment or prevention of a wide variety of
XX diseases including neurological disorders (e.g., Alzheimer's disease,
XX depression, diabetic neuropathy, Parkinson's disease or schizophrenia);
XX disorders of the adrenal gland; disorders of the colon or intestine
XX (e.g., Crohn's disease, diarrhoea, food poisoning or irritable bowel
XX syndrome); cardiovascular disorders (e.g., angina, cardiac arrhythmia or
XX myocardial infarction); muscular disorders; blood disorders (e.g.,
XX anaemia or leukaemia); immune disorders (e.g., osteoarthritis, rheumatoid
XX AIDS); bone and joint disorders (e.g., osteoporosis); metabolic or nutritive disorders (e.g.,
XX arthritis, gout or osteoporosis); metabolic or nutritive deficiency-related
XX diseases; enzyme deficiency-related diseases or vitamin deficiency-related
XX diseases); and disorders of the kidney, liver, lung, breast, ovary,
XX uterus, prostate, testis, skin, stomach, pancreas, spleen, thymus and
XX thyroid (e.g., cancers). The present sequence represents a GPCR of the
XX invention. Note: The full sequence data for this patent did not form part
```

CC of the printed specification; those sequences not shown were obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 2481 AA;

```
Query Match      4.1%; Score 101; DB 8; Length 2481;
Best Local Similarity 18.8%; Pred. No. 30; Mismatches 150; Indels 140; Gaps 19;
Matches 81; Conservative 61;

Qy 69 TISTKYNVSKQPVKNTLNTTEFPQYIILAGPIONYSITYLMWDF-----112
Db 1757 SLHTSFNIQVSP--SLT-----SFKSASGPTKVKVKTTCFSSNTRKMTSLEKTSLTN 1808

Qy 113 YSTQLRPAKY-----VYQYNHTAKTITFR--PPCGRVPSMTCLSEM 154
Db 1809 YATSLNTPVSPYPPWTPSSATLPSLTSFYVSPHSTEABISTPKISPPPTSQMVPEPVLGTR 1868

Qy 155 LNVSKRNDTGQCGCNFTTFNPFNVPRWTKLYVG-----PTKVNVDSTQIY 203
Db 1869 M-----TSSNTQPLMTSMNIPITAEBSQFPISITINVPITSNEMETETLH 1912

Qy 204 FL-----GLTALLLR--YAQRNCTHSFVLVN-AMSRNLFVRVPKYINGTKLKNTMRK 251
Db 1913 LVPGLSTFTTASQTGLVSKDVMAMSSIPMSGILPNHGLSEN-----PSLSTSLRA 1962

Qy 252 LKRQKAPVKEQEKAKK-TOSTTTPYF--SYTTSAAALNVTTNVTYSITTA-----299
Db 1963 ITSTLADVKKHTFEKWTTSVTPGTTLPSTLSCATSGSVISKSPILTWLLSSLPSPGPPATV 2022

Qy 300 --ARRVSTSTIAYRPSDSFMKSIATQRLDIAWYITLYRQNPFPCEPSNRRTAVSEFM 357
Db 2023 SNAPHVMTSSTVEVSKSTFLTSDMIS-----AHPFTNLTLPSAT 2062

Qy 358 KNTHLIRNETPYTIYG-----TLDMSLYYNETMFV-----ENKTASDSNKTPTS 404
Db 2063 MST-ILTRTPTPLGGITGPTSLPMSINVTDDIVITSTHPEASSKRTITANPRVSH 2121

Qy 405 PSMGFORTPIDP 416
Db 2122 PS-SFSRKTWSP 2132
```

Search completed: December 7, 2004, 15:17:13  
Job time : 160 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:03:50 ; Search time 200 Seconds  
(without alignments)  
1337.746 Million cell updates/sec

Title: US-09-942-146A-1  
Perfect score: 2451  
Sequence: 1 MGRKEMVRDVPKMFVLISI.....PEHRRVNLSTNSLWMLQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_treml.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2416.5	98.6	466	1 UL74_HCMVA	P16750 human cytom
2	2416.5	98.6	466	2 Q7M6L8	Q7m6l8 human cytom
3	2416.5	98.6	466	2 AAN40055	Aan40055 human cyt
4	2410.5	98.3	466	2 Q8AZ28	Q8az28 human cytom
5	2406.5	98.2	466	2 Q8BCU8	Q8bcu8 human cytom
6	2397.5	97.8	466	2 Q8AZ36	Q8az36 human cytom
7	2397.5	97.8	466	2 Q8BCU2	Q8bcu2 human cytom
8	2068	84.4	463	2 Q8AZ32	Q8az32 human cytom
9	2065	84.3	463	2 Q8BCV9	Q8bcv9 human cytom
10	2065	84.3	463	2 Q8BCV2	Q8bcv2 human cytom
11	2064	84.2	463	2 Q8AZ39	Q8az39 human cytom
12	2063.5	84.2	464	2 Q8BCU3	Q8bcu3 human cytom
13	2051.5	83.7	464	2 Q8BCV8	Q8bcv8 human cytom
14	2051.5	83.7	464	2 AAS48965	Aas48965 human cyt
15	2050.5	83.7	462	2 Q8BCU0	Q8bcu0 human cytom
16	2045.5	83.5	462	2 Q8BCU6	Q8bcu6 human cytom
17	1942	79.2	465	2 Q8AYW0	Q8ayw0 human cytom
18	1936	79.0	465	2 Q8BCU9	Q8bcu9 human cytom
19	1899	77.5	462	2 Q8AYZ3	Q8ayz3 human cytom
20	1894.5	77.3	462	2 Q8BCV5	Q8bcv5 human cytom
21	1870.5	76.3	472	2 Q8BCV1	Q8bcv1 human cytom
22	1870.5	76.3	472	2 AAR31626	Aar31626 human cyt
23	1748.5	71.3	464	2 Q8BCU1	Q8bcu1 human cytom
24	1740.5	71.0	464	2 Q8BCU4	Q8bcu4 human cytom
25	1737.5	70.9	464	2 Q8AYV5	Q8ayv5 human cytom
26	1735.5	70.8	464	2 Q8AYZ5	Q8ayz5 human cytom
27	1730	70.6	457	2 Q8BCV7	Q8bcv7 human cytom
28	1730	70.6	457	2 Q8BCV0	Q8bcv0 human cytom
29	977.5	39.9	461	2 Q8QS25	Q8qs25 chimpanzee
30	772	31.5	389	2 Q7TFN1	Q7tfn1 thesus cyto
31	717.5	29.3	144	2 Q7T9B3	Q7t9b3 human cytom

32	717.5	29.3	144	2	Q7T9B4	Q7t9b4 human cytom
33	717.5	29.3	144	2	Q7T9B5	Q7t9b5 human cytom
34	717.5	29.3	144	2	Q7T9B6	Q7t9b6 human cytom
35	717.5	29.3	144	2	Q7T9C3	Q7t9c3 human cytom
36	717.5	29.3	144	2	Q7T9C4	Q7t9c4 human cytom
37	717.5	29.3	144	2	Q7T9C5	Q7t9c5 human cytom
38	717.5	29.3	144	2	Q7T9D2	Q7t9d2 human cytom
39	717.5	29.3	144	2	Q7T9D4	Q7t9d4 human cytom
40	717.5	29.3	144	2	Q7T9D5	Q7t9d5 human cytom
41	706.5	28.8	144	2	Q7T9B2	Q7t9b2 human cytom
42	678.5	27.7	144	2	Q7T9C1	Q7t9c1 human cytom
43	497	20.3	144	2	Q7T9B8	Q7t9b8 human cytom
44	497	20.3	144	2	Q7T9D0	Q7t9d0 human cytom
45	491	20.0	142	2	Q7T9B9	Q7t9b9 human cytom

## ALIGNMENTS

### RESULT 1

ID	UL74_HCMVA	STANDARD;	PRT;	466 AA.
AC	P16750;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Glycoprotein-UL74-precursor;			
GN	Name=UL74;			
OS	Human cytomegalovirus (strain AD169).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Betaherpesvirinae; Cytomegalovirus.			
OX	NCBI_TaxID=10360;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE:90269039; PubMed:2161319;			
RA	Chen M.S., Banker A.T., Beck S., Bohni R., Brown C.M., Cerny R.,			
RA	Horsnell T., Hutchison C.A. III, Kourazides T., Martignetti J.A.,			
RA	Reddick E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;			
RT	"Analysis of the protein-coding content of the sequence of human			
RT	cytomegalovirus strain AD169.";			
RT	Curr. Top. Microbiol. Immunol. 154:125-169(1990).			
CC	-1- SIMILARITY: Belongs to the herpesviruses/UL7 family			

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL: X17403; CAA35389.1; -				
DR	PIR; S09837; S09837;				
KW	Glycoprotein; Signal;				
FT	CHAIN 1 30				
FT	SIGNAL 31 466				Potential.
FT	CARBOHYD 75 75				Glycoprotein UL74.
FT	CARBOHYD 83 83				N-linked (GlcNAc. .)
FT	CARBOHYD 87 87				N-linked (GlcNAc. .)
FT	CARBOHYD 103 103				N-linked (GlcNAc. .)
FT	CARBOHYD 130 130				N-linked (GlcNAc. .)
FT	CARBOHYD 157 157				N-linked (GlcNAc. .)
FT	CARBOHYD 162 162				N-linked (GlcNAc. .)
FT	CARBOHYD 171 171				N-linked (GlcNAc. .)
FT	CARBOHYD 219 219				N-linked (GlcNAc. .)
FT	CARBOHYD 242 242				N-linked (GlcNAc. .)
FT	CARBOHYD 288 288				N-linked (GlcNAc. .)
FT	CARBOHYD 292 292				N-linked (GlcNAc. .)
FT	CARBOHYD 350 350				N-linked (GlcNAc. .)
FT	CARBOHYD 385 385				N-linked (GlcNAc. .)
FT	CARBOHYD 392 392				N-linked (GlcNAc. .)
FT	CARBOHYD 399 399				N-linked (GlcNAc. .)
FT	CARBOHYD 433 433				N-linked (GlcNAc. .)



```
Db 121 PAKYVSYQNHAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVPWNTKLYVGPVKVNDVSDQTIYFLGLTALLRYAQRNCTHSHFYLNVAMSRNLFRVPKY 239
Db 181 NVPWNTKLYVGPVKVNDVSDQTIYFLGLTALLRYAQRNCTHSHFYLNVAMSRNLFRVPKY 240
Qy 240 INGTLKNTMKLRKQAPVKEQPEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 299
Db 241 INGTLKNTMKLRKQAPVKEQPEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 300
Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 360
Qy 360 THVLIRNETPTIYIGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGTFTFDPLWD 419
Db 361 THVLIRNETPTIYIGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGTFTFDPLWD 420
Qy 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 466
```

## RESULT 4

```
Q8AZ28 ID Q8AZ28 PRELIMINARY; PRT; 466 AA.
AC Q8AZ28;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=298, SW1, and SW3;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531320; AAN40044.1; -
DR EMBL; AF531325; AAN40049.1; -
DR EMBL; AF531337; AAN40061.1; -
DR EMBL; AF531346; AAN40070.1; -
SQ SEQUENCE 466 AA; 54200 MW; 56581B75919572FE CRC64;
```

Query Match 98.3%; Score 2410.5; DB 2; Length 466;

Best Local Similarity 98.7%; Pred. No. 1.3e-177;

Matches 460; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
Qy 1 MGRKEMVDRVDPKMFVLISISFLVSNCKVMSKALYNRPWGLVLSKIKYKLDQKL 60
Db 1 MGRKEMVDRVDPKMFVLISISFLVSNCKVMSKALYNRPWGLVLSKIKYKLDQKL 60
Qy 61 EILRQLETTISTKYNVSKQPVKNLTMTN-TEPQYVILAGPIQNTSYTLWDFDYSTQLRK 119
Db 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEPQYVILAGPIQNTSYTLWDFDYSTQLRK 120
Qy 120 PAKYVSYQNHAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKYVSYQNHAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVPWNTKLYVGPVKVNDVSDQTIYFLGLTALLRYAQRNCTHSHFYLNVAMSRNLFRVPKY 239
Db 181 NVPWNTKLYVGPVKVNDVSDQTIYFLGLTALLRYAQRNCTHSHFYLNVAMSRNLFRVPKY 240
Qy 240 INGTLKNTMKLRKQAPVKEQPEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 299
Db 241 INGTLKNTMKLRKQAPVKEQPEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 300
```

```
Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 360
Qy 360 THVLIRNETPTIYIGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGTFTFDPLWD 419
Db 361 THVLIRNETPTIYIGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGTFTFDPLWD 420
Qy 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 466
```

## RESULT 5

```
Q8BCU8 ID Q8BCU8 PRELIMINARY; PRT; 466 AA.
AC Q8BCU8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=851;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531330; AAN40054.1; -
SQ SEQUENCE 466 AA; 54230 MW; 56581B75919562PF CRC64;
```

Query Match 98.2%; Score 2406.5; DB 2; Length 466;

Best Local Similarity 98.5%; Pred. No. 2.6e-177;

Matches 459; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```
Qy 1 MGRKEMVDRVDPKMFVLISISFLVSNCKVMSKALYNRPWGLVLSKIKYKLDQKL 60
Db 1 MGRKEMVDRVDPKMFVLISISFLVSNCKVMSKALYNRPWGLVLSKIKYKLDQKL 60
Qy 61 EILRQLETTISTKYNVSKQPVKNLTMTN-TEPQYVILAGPIQNTSYTLWDFDYSTQLRK 119
Db 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEPQYVILAGPIQNTSYTLWDFDYSTQLRK 120
Qy 120 PAKYVSYQNHAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKYVSYQNHAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVPWNTKLYVGPVKVNDVSDQTIYFLGLTALLRYAQRNCTHSHFYLNVAMSRNLFRVPKY 239
Db 181 NVPWNTKLYVGPVKVNDVSDQTIYFLGLTALLRYAQRNCTHSHFYLNVAMSRNLFRVPKY 240
Qy 240 INGTLKNTMKLRKQAPVKEQPEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 299
Db 241 INGTLKNTMKLRKQAPVKEQPEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 300
Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 360
Qy 360 THVLIRNETPTIYIGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGTFTFDPLWD 419
Db 361 THVLIRNETPTIYIGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGTFTFDPLWD 420
Qy 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 466
```

```
RESULT 6
Q8AZ36 PRELIMINARY; PRT; 466 AA.
AC Q8AZ36;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4088, and SW1762;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531322; AAN40046.1; -.
DR EMBL; AF531344; AAN40068.1; -.
SQ SEQUENCE 466 AA; 54215 MW; 4DC855844E3B65A0 CRC64;

Query Match 97.8%; Score 2397.5; DB 2; Length 466;
Best Local Similarity 98.3%; Pred. No. 1.3e-176;
Matches 458; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MGRKMMVRDVPKMFVLISISFLVSVFNCVKMSKALYNRPWRGLVLSKIGYKLDQLK 60
DB 1 MGRKMMVRDVPKMFVLISISFLVSVFNCVKMSKALYNRPWRGLVLSKIGYKLDQLK 60
QY 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNYSIYILWDFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQPVKNLTWNTEFPQYIILAGPIQNYSIYILWDFYSTQLRK 120
QY 120 PAKYVSYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 PAKYVSYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
QY 180 NVRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
DB 181 NVRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
QY 240 INGTLKNTMKLRKQAPVKEQFEKAKKTQSTITTPYFSYTTSAALNVTTNVTYSITTA 299
DB 241 INGTLKNTMKLRKQAPVKEQFEKAKKTQSTITTPYFSYTTSAALNVTTNVTYSITTA 300
QY 300 ARRVSTSTIAYRPDSSFMKSIATQRLDRLATWVYTLRYQNPFCEPSRNRRTAVSEPMKN 359
DB 301 ARRVSTSTIAYRPDSSFMKSIATQRLDRLATWVYTLRYQNPFCEPSRNRRTAVSEPMKN 360
QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSGMFGORTFIDPLWD 419
DB 361 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSGMFGORTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSLRSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSLRSPYVNLTPPEHRRVAVNLSTNSLWMLQ 466

RESULT 7
Q8BCU2 PRELIMINARY; PRT; 466 AA.
AC Q8BCU2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=279, and SW2;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531319; AAN40043.1; -.
DR EMBL; AF531345; AAN40069.1; -.
SQ SEQUENCE 463 AA; 53871 MW; 51310C0D1D239A5A CRC64;

Query Match 97.8%; Score 2397.5; DB 2; Length 466;
Best Local Similarity 98.3%; Pred. No. 1.3e-176;
Matches 458; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MGRKMMVRDVPKMFVLISISFLVSVFNCVKMSKALYNRPWRGLVLSKIGYKLDQLK 60
DB 1 MGRKMMVRDVPKMFVLISISFLVSVFNCVKMSKALYNRPWRGLVLSKIGYKLDQLK 60
QY 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNYSIYILWDFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQPVKNLTWNTEFPQYIILAGPIQNYSIYILWDFYSTQLRK 120
QY 120 PAKYVSYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 PAKYVSYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
QY 180 NVRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
DB 181 NVRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
QY 240 INGTLKNTMKLRKQAPVKEQFEKAKKTQSTITTPYFSYTTSAALNVTTNVTYSITTA 299
DB 241 INGTLKNTMKLRKQAPVKEQFEKAKKTQSTITTPYFSYTTSAALNVTTNVTYSITTA 300
QY 300 ARRVSTSTIAYRPDSSFMKSIATQRLDRLATWVYTLRYQNPFCEPSRNRRTAVSEPMKN 359
DB 301 ARRVSTSTIAYRPDSSFMKSIATQRLDRLATWVYTLRYQNPFCEPSRNRRTAVSEPMKN 360
QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSGMFGORTFIDPLWD 419
DB 361 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSGMFGORTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSLRSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSLRSPYVNLTPPEHRRVAVNLSTNSLWMLQ 466

RESULT 8
Q8AZ32 PRELIMINARY; PRT; 463 AA.
AC Q8AZ32;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=279, and SW2;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531319; AAN40043.1; -.
DR EMBL; AF531345; AAN40069.1; -.
SQ SEQUENCE 463 AA; 53871 MW; 51310C0D1D239A5A CRC64;
```

Query Match	84.4%	Score 2068	DB 2	Length 463
Best Local Similarity	85.4%	Pred. No. 3.5e-151		
Matches 399	Conservative 22	Mismatches 40	Indels 6	Gaps 4
QY 1	MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL	60		
DB 1	MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL	60		
QY 61	EILRLQLETTISTKY-NVSKOPVKNLWN-TEFPQYIILAGPIQNSYITILWDFYSTQLR	118		
DB 60	EILRLQLETTISTKY-NVSKOPVKNLWN-TEFPQYIILAGPIQNSYITILWDFYSTQLR	119		
QY 119	KPAKYVYSQNHAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGCNFTTFNPMF	178		
DB 120	KPAKYVSEYNHTAKTITFRPPSCGTVPSTCLSEMLNVSKRNDTGEQCGCNFTTFNPMF	179		
QY 179	FNVPWNTKLYVGPVKNVDSQTIYFGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK	238		
DB 180	FNVPWNTKLYVGSKKVNVDSQTIYFGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK	239		
QY 239	YINGTKLNTMKRLKRQAPVKEQFEKKAKKTQSTTTTPYFSYITSAALNVTTNTYSITT	298		
DB 240	YINGTKLNTMKRLKRQAPVKEQSEKSKSQSTTTPYSPYTTSTALNVTTNTYSITT	299		
QY 299	AARVSTSTIAYRPDSSFMKSIMATQRLDALTWYTTILRYQRNCFPCPSRNRRTAVSBFMK	358		
DB 300	TARRVSTSTIAYRPDSSFMKSIMTQLRLDALTWYTTILRYQRNCFPCSSRNRRTAVSBFMK	359		
QY 359	NTHVLINETPYTITGLDSSLYNMTFVENKTASDNKTTTSPSMGFQRTFIDPLW	418		
DB 360	NTHVLINETPYTITGLDSSLYNMTFVENKTASE--TTPSPSTGFQRTFIDPLW	416		
QY 419	DYLDLSLLFLDEIRNFSRSPYVNLTPPEHRRVNLSTNSLWMLQ	465		
DB 417	DYLDLSLLFLDEIRNFSQSTPYGNLTPEHRRVNLSTNSLWMLQ	463		
RESULT 9				
Q8BCT9	PRELIMINARY;	PRT;	463 AA.	
ID	Q8BCT9			
AC	Q8BCT9			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	UL74 protein.			
OS	Human cytomegalovirus.			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Betaherpesvirinae; Cytomegalovirus.			
OX	NCBI_TaxID=10359;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SW990;			
RX	MEDLINE=22255414; PubMed=12368327;			
RT	Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;			
RT	"The genes encoding the gCII complex of human cytomegalovirus exist			
RT	in highly diverse combinations in clinical isolates.";			
DR	J. Virol. 76:10841-10848(2002).			
DR	EMBL; AF531354; AAN40078.1; -.			
SQ	SEQUENCE 463 AA; 53985 MW; B86783A05FB9424A CEC64;			
Query Match	84.3%	Score 2065	DB 2	Length 463
Best Local Similarity	85.2%	Pred. No. 5.9e-151		
Matches 398	Conservative 23	Mismatches 40	Indels 6	Gaps 4
QY 1	MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL	60		
DB 1	MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL	60		
QY 61	EILRLQLETTISTKY-NVSKOPVKNLWN-TEFPQYIILAGPIQNSYITILWDFYSTQLR	118		
DB 60	EILRLQLETTISTKY-NVSKOPVKNLWN-TEFPQYIILAGPIQNSYITILWDFYSTQLR	119		
QY 119	KPAKYVYSQNHAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGCNFTTFNPMF	178		

Db	120	KPAKYVFSEYNHTAKTITFRPPSCGTVPSMTCLSEMNLVSKRNDTGEGCGNFTTFNPMF	179
Qy	179	FNVPRWNTKLYVGPTKVNVDSQTIFYFLGLTALLRYAQRNCCTHSPYLNAMSRNLFVRPK	238
Db	180	FNVPRWNTKLYVGSKKVNVDSQTIFYFLGLTALLRYAQRNCCTHSPYLNAMSRNLFVRPK	239
Qy	239	YINGTKLKNTRMKLRKQAOPVKGEQFEKKAKTQSTTTTPYSYTTSAALNVTNTVTSITT	298
Db	240	YINGTKLKNTRMKLRKQAOPVKGEQEKSKKSQSSTTTPSPYTTSTALNVTTNATYSVTT	299
Qy	299	AARVSTSTIAYRPDSSFMKSIMATQLRDLATWYTTLIRYQNPFCEPSNRRTAVSEPMK	358
Db	300	TABRVSTSTIAYRPDSSFMKSIMTQURDLATWYTTLIRYQNPFCESSRNRTAVSEPMK	359
Qy	359	NTHVLIRNETPYTYIGTLDMSLLYYNMTMFVENKTASDNKTTPSPMGFORPIDPLM	418
Db	360	NTHVLIRNETPYTYIGTLDMSLLYYNMTMFVENTASE---TPTSSTGFQRFIDPLM	416
Qy	419	DYLOSLFLDEIRNFSLRSTPYVNLTPPEHRRANLSTNSLWMLQ	465
Db	417	DYLOSLFLDEIRNFSLQSPTYGNLTPEHRRANLSTLSLWMLQ	463
RESULT 10			
Q8BCV2	ID	Q8BCV2 PRELIMINARY; PRT; 463 AA.	
AC	Q8BCV2;		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	UL74 protein.		
OS	Human cytomegalovirus.		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Betaherpesvirinae; Cytomegalovirus.		
OX	NCBI_TaxID=10359;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	STRAIN=122;		
RX	MEDLINE=22255414; PubMed=12368327;		
RA	Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;		
RT	"The genes encoding the gCII complex of human cytomegalovirus exist		
RT	in highly diverse combinations in clinical isolates."		
RL	J. Virol. 76:10841-10848(2002).		
DR	EMBL; AF531316; AAN4004.1; -.		
SQ	SEQUENCE 463 AA; 53857 MW; 55A7D4157FD32424 CRC64;		
Query Match			
Best Local Similarity 84.3%; Score 2065; DB 2; Length 463;			
Matches 399; Conservative 22; Mismatches 40; Indels 6; Gaps 4			
Qy	1	MGRKEHMVRDPKMFVLISIFLLSVSPINCKVMKALYNRPWRGLVLSKGICYKLDDCLK	60
Db	1	MGRKGM-RGVFLFFLMSTITLFLSFPCINRAAVRLSGVRWWSKVLTGQRUDKFKL	59
Qy	61	EILRLQLETTISTKY-NVSKQPVKNLTWN--TEFPQYIILAGPIQNYISITYLWFDFYSTOLR	118
Db	60	EILKQLEKDITTKYFNWTRQHINKULTNMNTEFPFYIILAGPIQNNSVTYLVWDFVSTQLR	119
Qy	119	KPAKYVSQNHHTAKTITFRPPPCGRVPSTMCLSEMNLVSKRNDTGEGCGNFTTFNPMF	178
Db	120	KPAKYVFSEYNHTAKTITFRPPSCGTVPSMTCLSEMNLVSKRNDTGEGCGNFTTFNPMF	179
Qy	179	FNVPRWNTKLYVGPTKVNVDSQTIFYFLGLTALLRYAQRNCCTHSPYLNAMSRNLFVRPK	238
Db	180	FNVPRWNTKLYVGSKKVNVDSQTIFYFLGLTALLRYAQRNCCTHSPYLNAMSRNLFVRPK	239
Qy	239	YINGTKLKNTRMKLRKQAOPVKGEQFEKKAKTQSTTTTPYSYTTSAALNVTNTVTSITT	298
Db	240	YINGTKLKNTRMKLRKQAOPVKGEQEKSKKSQSSTTTPSPYTTSTALNVTTNATYSVTT	299
Qy	299	AARVSTSTIAYRPDSSFMKSIMATQLRDLATWYTTLIRYQNPFCEPSNRRTAVSEPMK	358

Db 300 TARRVSTSTIAYRPDSSFMKSIIMTQLRDLATWVYTLRYRQNPFCSSRNRTAVSEFMK 359  
Qy 359 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418  
Db 360 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASE--TTPTSPSTGQRTFIDPLW 416  
Qy 419 DYLDLFLDLBEIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465  
Db 417 DYLDLFLDLBEIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 463

RESULT 11  
Q8AZ39  
ID Q8AZ39 PRELIMINARY; PRT; 463 AA.  
AC Q8AZ39  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE UL74 protein.  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=650, and DM7;  
RX MEDLINE=22255414; PubMed=12368327;  
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;  
RT "The genes encoding the gCIII complex of human cytomegalovirus exist  
in highly diverse combinations in clinical isolates.";  
RL J. Virol. 76:10841-10848(2002).  
DR EMBL; AF531326; AAN40050.1; -  
DR EMBL; AF531334; AAN40058.1; -  
SQ SEQUENCE 463 AA; 53901 MW; 403100AABDF4412A CRC64;

Query Match 84.2%; Score 2064; DB 2; Length 463;  
Best Local Similarity 85.2%; Pred. No. 7e-151;  
Matches 398; Conservative 22; Mismatches 41; Indels 6; Gaps 4;  
Qy 1 MGRKEMVVDVPMFVLISIFLVSFINCKVMSKALNRPWRGLVLSKIGYKLDOLKL 60  
Db 1 MGRKEM-RGVNLFPLMSLTLFSLFNCRAAIVLSGVYSGKVLSTIGKQKLDKFKL 59  
Qy 61 EILRQLETTISTKY-NVSKQPVKQLTNW-TEFPQYIILAGPIQNSITYLWDFYSTQLR 118  
Db 60 EILQLEKDIYTKYFNWTRQHKNLWNMTFEPYIILAGPIQNSITYLWDFYSTQLR 119  
Qy 119 KPAKYVSYQNHNTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTFNPMP 178  
Db 120 KPAKYVSEYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTFNPMP 179  
Qy 179 FNVPRWNTKLVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSEFYLNVMSRNLFRVVK 238  
Db 180 FNVPRWNTKLVGSKVNVDSQTIYFLGLTALLRYAQRNCTHSEFYLNVMSRNLFRVVK 239  
Qy 239 YINGTKLKNWTKLRKQAPVKEQFEKAKTKQSTTTPYFSYTTSAALNVNTNYSITT 298  
Db 240 YINGTKLKNWTKLRKQAPVKEQSEKSKSQSTTTPYFSYTTSTALNVNTNYSITT 299  
Qy 299 AARVSTSTIAYRPDSSFMKSIIMATQRLDNLATWVYTLRYRQNPFCSSRNRTAVSEFMK 358  
Db 300 TTRVSTSTIAYRPDSSFMKSIIMTQLRDLATWVYTLRYRQNPFCSSRNRTAVSEFMK 359  
Qy 359 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418  
Db 360 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASE--TTPTSPSTGQRTFIDPLW 416  
Qy 419 DYLDLFLDLBEIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465  
Db 417 DYLDLFLDLBEIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 463

RESULT 12

Q8BCU3  
ID Q8BCU3 PRELIMINARY; PRT; 464 AA.  
AC Q8BCU3  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE UL74 protein.  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SW1324;  
RX MEDLINE=22255414; PubMed=12368327;  
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;  
RT "The genes encoding the gCIII complex of human cytomegalovirus exist  
in highly diverse combinations in clinical isolates.";  
RL J. Virol. 76:10841-10848(2002).  
DR EMBL; AF531340; AAN40064.1; -  
SQ SEQUENCE 464 AA; 54146 MW; A24A513BF112848B CRC64;

Query Match 84.2%; Score 2063.5; DB 2; Length 464;  
Best Local Similarity 84.8%; Pred. No. 7.7e-151;  
Matches 396; Conservative 25; Mismatches 41; Indels 5; Gaps 4;  
Qy 1 MGRKEMVVDVPMFVLISIFLVSFINCKVMSKALNRPWRGLVLSKIGYKLDOLKL 60  
Db 1 MGRKEDM-RIISKLFPIISITLVLLFSIINCKVVRPP--GRYWLGTVLSTIGKQKLDKFKL 57  
Qy 61 EILRQLETTISTKY-NVSKQPVKQLTNW-TEFPQYIILAGPIQNSITYLWDFYSTQLR 118  
Db 58 EILKQLEREPEPTYKFNWTRQHKNLWNMTQFPQYIILAGPIRNDSTIYLWDFYSTQLR 117  
Qy 119 KPAKYVSYQNHNTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTFNPMP 178  
Db 118 KPAKYVSYQNHNTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTFNPMP 177  
Qy 179 FNVPRWNTKLVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSEFYLNVMSRNLFRVVK 238  
Db 178 FNVPRWNTKLVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSEFYLNVMSRNLFRVVK 237  
Qy 239 YINGTKLKNWTKLRKQAPVKEQFEKAKTKQSTTTPYFSYTTSAALNVNTNYSITT 298  
Db 238 YINGTKLKNWTKLRKQAPVKEQLEKTKKSQSTTTPYFSYTTSTALNVNTNYSITT 297  
Qy 299 AARVSTSTIAYRPDSSFMKSIIMATQRLDNLATWVYTLRYRQNPFCSSRNRTAVSEFMK 358  
Db 298 SAKRIPTSTIAYRPDSSFMKSIIMATQRLDNLATWVYTLRYRNEPFCPPDRNRTAVSEFMK 357  
Qy 359 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418  
Db 358 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTTPTSPSTGQRTFIDPLW 417  
Qy 419 DYLDLFLDLBEIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465  
Db 418 DYLDLFLDLKIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 464

RESULT 13  
Q8BCT8  
ID Q8BCT8 PRELIMINARY; PRT; 464 AA.  
AC Q8BCT8  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE UL74 protein.  
GN Name=UL74; ORFNames=HHV5gp069;  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10359;  
RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=Toledo;  
 RX MEDLINE=2255414; PubMed=12369327;  
 RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;  
 RT "The genes encoding the gCIII complex of human cytomegalovirus exist  
 in highly diverse combinations in clinical isolates.";  
 RL J. Virol. 76:10841-10848(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Toledo;  
 RA Brondke H., Schmitz B., Shenk T., Doerfler W.;  
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF531355; AAN40073.1; -;  
 DR EMBL; AY486475; AAS48965.1; -;  
 SQ SEQUENCE 464 AA; 54020 MW; 7751A6829CE6E3E7 CRC64;  
 Query Match 83.7%; Score 2051.5; DB 2; Length 464;  
 Best Local Similarity 84.4%; Pred. No. 6.5e-150;  
 Matches 394; Conservative 25; Mismatches 43; Indels 5; Gaps 4;  
 Qy 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWGLVLSKIGYKLDQLKL 60  
 Db 1 MGRKGM-RSISKLFPIISLTVLLFSIINCKVVRPP--GRYMLGTVLSTIGKQKLDKPKL 57  
 Qy 61 EILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLR 118  
 Db 58 EILQLEREPYTYFNMTQHVKNLTNMTQFPQYIILAGPIRNDSTIYLWFDYFYSTQLR 117  
 Qy 119 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMF 178  
 Db 118 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMF 177  
 Qy 179 FNVRMNTKLYVGTQKVNDSQTIYFLGLTALLRYAQRNCTHSEFYLVNMSRNLFRVPK 238  
 Db 178 FNVRMNTKLYVGTQKVNDSQTIYFLGLTALLRYAQRNCTHSEFYLVNMSRNLFRVPK 237  
 Qy 239 YINGTKLNTMRKLRKQAPVKEQFEKAKTKTQSTTTTPYSYTTSAALNVTNTYISITT 298  
 Db 238 YINGTKLNTMRKLRKQAPVKEQFEKAKTKTQSTTTTPYSYTTSAALNVTNTYISITT 297  
 Qy 299 AARVSTSTIAYRPDSSFMKSIMATQRLDWTATVYTLRYQRNPFCEPSNRRTAVSEFMK 358  
 Db 298 SAKRIPSTIAYRPDSSFMKSIMATQRLDWTATVYTLRYQRNPFCEPSNRRTAVSEFMK 357  
 Qy 359 NTHVLIRNETPYTYIGTLDSSLYNETMFVENKTASDNKTTTSPSMGQRTFIDPLW 418  
 Db 358 NTHVLIRNETPYTYIGTLDSSLYNETMFVENKTASDNKTTTSPSMGQRTFIDPLW 417  
 Qy 419 DYLDLSLFLDEIRNFSRSPYVNLTPPEHRAVNLSTSLMWLQ 465  
 Db 418 DYLDLSLFLDKIRNFSQLPAYGNLTPPEHRAVNLSTSLMWWSQ 464  
 RESULT 14  
 AAS48965 PRELIMINARY; PRT; 464 AA.  
 AC AAS48965;  
 DT 10-MAR-2004 (TrEMBLrel. 27, Created)  
 DT 10-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 10-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
 DE UL74.  
 GN UL74 OR HHV5GP069.  
 OS Human cytomegalovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OX NCBI\_TaxID=10359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Toledo;  
 RA Brondke H., Schmitz B., Shenk T., Doerfler W.;  
 RT "Differences in the Nucleotide Sequences between the AD169 and Toledo  
 Strains of Human Cytomegalovirus.";  
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY486475; AAS48965.1; -;  
 SQ SEQUENCE 464 AA; 54020 MW; 7751A6829CE6E3E7 CRC64;  
 Query Match 83.7%; Score 2051.5; DB 2; Length 464;  
 Best Local Similarity 84.4%; Pred. No. 6.5e-150;  
 Matches 394; Conservative 25; Mismatches 43; Indels 5; Gaps 4;  
 Qy 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWGLVLSKIGYKLDQLKL 60  
 Db 1 MGRKGM-RSISKLFPIISLTVLLFSIINCKVVRPP--GRYMLGTVLSTIGKQKLDKPKL 57  
 Qy 61 EILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLR 118  
 Db 58 EILQLEREPYTYFNMTQHVKNLTNMTQFPQYIILAGPIRNDSTIYLWFDYFYSTQLR 117  
 Qy 119 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMF 178  
 Db 118 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMF 177  
 Qy 179 FNVRMNTKLYVGTQKVNDSQTIYFLGLTALLRYAQRNCTHSEFYLVNMSRNLFRVPK 238  
 Db 178 FNVRMNTKLYVGTQKVNDSQTIYFLGLTALLRYAQRNCTHSEFYLVNMSRNLFRVPK 237  
 Qy 239 YINGTKLNTMRKLRKQAPVKEQFEKAKTKTQSTTTTPYSYTTSAALNVTNTYISITT 298  
 Db 238 YINGTKLNTMRKLRKQAPVKEQFEKAKTKTQSTTTTPYSYTTSAALNVTNTYISITT 297  
 Qy 299 AARVSTSTIAYRPDSSFMKSIMATQRLDWTATVYTLRYQRNPFCEPSNRRTAVSEFMK 358  
 Db 298 SAKRIPSTIAYRPDSSFMKSIMATQRLDWTATVYTLRYQRNPFCEPSNRRTAVSEFMK 357  
 Qy 359 NTHVLIRNETPYTYIGTLDSSLYNETMFVENKTASDNKTTTSPSMGQRTFIDPLW 418  
 Db 358 NTHVLIRNETPYTYIGTLDSSLYNETMFVENKTASDNKTTTSPSMGQRTFIDPLW 417  
 Qy 419 DYLDLSLFLDEIRNFSRSPYVNLTPPEHRAVNLSTSLMWLQ 465  
 Db 418 DYLDLSLFLDKIRNFSQLPAYGNLTPPEHRAVNLSTSLMWWSQ 464  
 RESULT 15  
 Q8BCU0 PRELIMINARY; PRT; 462 AA.  
 AC Q8BCU0;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE UL74 protein.  
 OS Human cytomegalovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OX NCBI\_TaxID=10359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SW490;  
 RX MEDLINE=2255414; PubMed=12369327;  
 RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;  
 RT "The genes encoding the gCIII complex of human cytomegalovirus exist  
 in highly diverse combinations in clinical isolates.";  
 RL J. Virol. 76:10841-10848(2002).  
 DR EMBL; AF531349; AAN40073.1; -;  
 SQ SEQUENCE 462 AA; 53738 MW; 8FAAAA0416E33D06 CRC64;  
 Query Match 83.7%; Score 2050.5; DB 2; Length 462;  
 Best Local Similarity 85.0%; Pred. No. 7.7e-150;  
 Matches 397; Conservative 23; Mismatches 40; Indels 7; Gaps 5;  
 Qy 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWGLVLSKIGYKLDQLKL 60  
 Db 1 MGRKGM-RGVNLFPLMSLTFLFPIINCRVAVLSVGRVSKVLSLSTIGKQKLDKPKL 59  
 Qy 61 EILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLR 118

Db 60 EILKQLEKDIYKYFNNTRQHIKNLWNMTFFPRYYILAGPIQNNNSVYILWDFDYSTQLR 119  
Qy 119 KPAKYVYSQVNHAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTNPMF 178  
Db 120 KPAKYVSEYNHAKTITFRPPSCGTVPSTCLSEMLNVSKRNDTGEQCGN-TTFNPMF 178  
Qy 179 FNVRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLVNAMSRLPRVPK 238  
Db 179 FNVRWNTKLYVGSKKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLVNAMSRLFRVPK 238  
Qy 239 YINGTKLNTMRKLKRQAPVKEQFEKKAKKTQSTTTYPFYSYTTSAALNVTNNVYSITT 298  
Db 239 YINGTKLNTMRKLKRQAPVKEQSEKSKSQSTTTYPSPYTTTALNVTNNATYSVTT 298  
Qy 299 AARVSTSTIAYRPDSSPFMSIMATQRLDLATWVYTTILRYRONPFCPSRNRNAVSEFMK 358  
Db 299 TARRISTSTIAYRPDSSPFMSIMTTQLRDLATWVYTTILRYRONPFCSSSRNRNAVSEFMK 358  
Qy 359 NTHVLIIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTPTSPSMGFQRTFIDPLW 418  
Db 359 NTHVLIIRNETPYTYIGTLDMSLYNETMFVENKTASE--TTFPSSTGFQRTFIDPLW 415  
Qy 419 DYLDLFLFLDEIRNFSLRSPYVNLTPPEHRRAVNLSLWMLQ 465  
Db 416 DYLDLFLFLDEIRNFSLQSPYGNLTPEHRRAVNLSLWMLQ 462

Search completed: December 7, 2004, 15:20:35  
Job time : 202 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:12:05 ; Search time 39 Seconds  
(without alignments)  
790.715 Million cell updates/sec

Title: US-09-942-146a-1  
Perfect score: 2451  
Sequence: 1 MGRKEMVRDVPKMFVLIS1.....PEHRAVNLSTNSLWNLQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2419	98.7	465	4	US-09-627-986-1
2	107.5	4.4	1168	1	US-08-620-717A-9
3	106.5	4.3	1167	1	US-08-485-568A-6
4	106.5	4.3	1167	2	US-08-590-554A-6
5	106.5	4.3	1167	2	US-09-184-223-6
6	105	4.3	452	4	US-09-543-681A-7874
7	101.5	4.1	351	4	US-09-248-796A-23675
8	101	4.1	476	4	US-09-248-796A-18994
9	99.5	4.1	978	2	US-08-415-593-43
10	97.5	4.0	1174	4	US-09-538-092-447
11	96.5	3.9	347	3	US-09-097-889-14
12	96.5	3.9	347	4	US-09-098-079-14
13	95.5	3.9	651	4	US-09-107-532A-4902
14	95.5	3.9	651	4	US-09-134-000C-5699
15	95	3.9	1065	4	US-09-538-092-84
16	95	3.9	2870	4	US-09-479-467A-15
17	95	3.9	3178	4	US-09-479-467A-15
18	94.5	3.9	354	4	US-09-248-796A-15271
19	94.5	3.9	383	4	US-09-248-796A-16746
20	94.5	3.9	515	4	US-09-248-796A-26713
21	94.5	3.9	705	4	US-09-248-796A-19971
22	94.5	3.9	844	4	US-09-991-258-19
23	94.5	3.9	2938	5	PCT-US94-00198-3
24	94	3.8	441	4	US-09-248-796A-23208
25	94	3.8	576	4	US-09-248-796A-20509
26	94	3.8	629	4	US-09-107-532A-6656
27	93.5	3.8	213	4	US-09-538-092-1255

28	93	3.8	541	4	US-09-248-796A-18318	Sequence 18318, A
29	92.5	3.8	537	4	US-09-248-796A-24443	Sequence 24443, A
30	92.5	3.8	1579	3	US-08-755-587-184	Sequence 184, App
31	92	3.8	288	4	US-09-216-393B-341	Sequence 341, App
32	92	3.8	288	4	US-09-216-393B-344	Sequence 344, App
33	91.5	3.7	250	4	US-09-248-796A-15562	Sequence 15562, A
34	91.5	3.7	354	3	US-09-307-925-8	Sequence 8, Appli
35	91.5	3.7	354	4	US-09-991-582B-8	Sequence 8, Appli
36	91.5	3.7	544	3	US-08-559-397A-30	Sequence 30, Appli
37	91	3.7	365	4	US-09-248-796A-15803	Sequence 15803, A
38	91	3.7	2186	2	US-08-822-445-2	Sequence 2, Appli
39	91	3.7	2186	3	US-09-396-540-2	Sequence 2, Appli
40	90.5	3.7	360	3	US-08-875-811-47	Sequence 47, Appli
41	90.5	3.7	721	4	US-09-248-796A-18800	Sequence 18800, A
42	90.5	3.7	797	4	US-09-134-000C-4997	Sequence 4997, Ap
43	90.5	3.7	1533	1	US-08-623-679-9	Sequence 9, Appli
44	90.5	3.7	1533	3	US-08-933-774-9	Sequence 9, Appli
45	90.5	3.7	1533	3	US-09-181-030-9	Sequence 9, Appli

## ALIGNMENTS

## RESULT 1

US-09-627-986-1

; Sequence 1, Application US/09627986

; Patent No. 6569616

; GENERAL INFORMATION:

; APPLICANT: Compton, Teresa

; APPLICANT: Huber, Mary T.

; TITLE OF INVENTION: HUMAN CYTOMEGALOVIRUS GLYCOPROTEIN O AS A NEW DRUG

; FILE REFERENCE: 960296.96579

; CURRENT APPLICATION NUMBER: US/09/627.986

; PRIOR FILING DATE: 2000-07-28

; PRIOR APPLICATION NUMBER: 60/146,180

; PRIOR FILING DATE: 1999-07-29

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; LENGTH: 465

; TYPE: PRT

; ORGANISM: Human cytomegalovirus

US-09-627-986-1

Query Match	98.7%	Score 2419;	DB 4;	Length 465;
Best Local Similarity	99.4%	Pred. No. 7.6e-235;		
Matches 463;	Conservative 0;	Mismatches 1;	Indels 2;	Gaps 2;
Qy	1	MGRKEMVRDVPKMFVLISISFLVSPINCKWMSKALYNRPWRGLVLSKIGKYKLDQLKL	60	
Db	1	MGRKEMVRDVPKMFVLISISFLVSPINCKWMSKALYNRPWRGLVLSKIGKYKLDQLKL	60	
Qy	61	EILROLETTISTKYNYSKQPKNLTWNTEPPQYILAGPIQNVISITVLPDPYSTOLRKP	120	
Db	61	EILROLETTISTKYNYSKQPKNLTWNTEPPQYILAGPIQNVISITVLPDPYSTOLRKP	120	
Qy	121	AKVYSQYNHTAKTIIPR-PPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF	179	
Db	121	AKVYSQYNHTAKTIIPRPPPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF	180	
Qy	180	NVPRWNTKLVGPTKVNVDSTQTYFLGLTALLRYAQRNCTHSPYLVNMSRNLFVRPKY	239	
Db	181	NVPRWNTKLVGPTKVNVDSTQTYFLGLTALLRYAQRNCTHSPYLVNMSRNLFVRPKY	240	
Qy	240	INGTKLKNTRKLRKQAPVKEQEKAKTKQSTTTTTFYSYTTSAALNVTNTVYSITTA	299	
Db	241	INGTKLKNTRKLRKQAPVKEQEKAKTKQSTTTTTFYSYTTSAALNVTNTVYSITTA	300	
Qy	300	ARRVSTSTIAYRPDSSPFMSIMATQRLDALTWVTTLYRQNPFCPSRNTAVSEFMKN	359	
Db	301	ARRVSTSTIAYRPDSSPFMSIMATQRLDALTWVTTLYRQNPFCPSRNTAVSEFMKN	359	

QY 360 THVLIRNETPTTYIGTLDSSLYNETMFVENKTASDNKTTPTSPSMGFORTFIDPLMD 419  
Db 360 THVLIRNETPTTYIGTLDSSLYNETMFVENKTASDNKTTPTSPSMGFORTFIDPLMD 419  
QY 420 YLDSLLFLDEIRNPSLRSPYVNLTPPEHRAVNLSTNSLWMLQ 465  
Db 420 YLDSLLFLDEIRNPSLRSPYVNLTPPEHRAVNLSTNSLWMLQ 465  
RESULT 2  
US-08-620-717A-9  
; Sequence 9, Application US/08620717A  
; Patent No. 5670365  
; GENERAL INFORMATION:  
; APPLICANT: Feitelson, Jerald S.  
; TITLE OF INVENTION: Identification of, and Uses For, Nematicidal  
; TITLE OF INVENTION: Bacillus thuringiensis Genes, Toxins, and Isolates  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/620,717A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/540,104  
; FILING DATE: 06-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA94.C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (352) 375-8100  
; TELEFAX: (352) 372-5800  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1168 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: 167p  
US-08-620-717A-9  
Query Match 4.4%; Score 107.5; DB 1; Length 1168;  
Best Local Similarity 20.9%; Pred. No. 0.22;  
Matches 97; Conservative 68; Mismatches 131; Indels 169; Gaps 26;  
QY 90 FPOY-ILAGPIQYISITYLWFDYSTQLRKPAPYQYQYNNHAKTITRPPPCGRVPSM 148  
Db 8 YPSYHNVLAPFLDS----FFDPF-VETFKDLKGAWEEF-----GKTGYM 48  
QY 149 TCSMLNYSKRNDSGEQCGNFTTNPMFFNPRWTKLYVGPVKVND-----SQ 200  
Db 49 DPLKQHLQIAW--DTSQNG-----TVDYIALTKASISLGLIPGADA 88  
QY 201 TIVFLGL-----TALLRYAQRNCTHSFV-----LVNMSRN--LFRVPKYNG- 242  
Db 89 VPPFINNVDFIPPKLFGSGSQQAQAFPELLIEKVELDVEDFENFNLNLLYDGM 148  
QY 243 ----TKLQNTMR-----KLKRKOAP-----VKEQPEKAKKTQSTTTPY 277

Db 149 QVALSHFQNDVQIAICQEQPGLMDQTPTACTTDDHLISVRESF-KDARTTITETALPH 207  
QY 278 FS-----YTTSAALNVTYTNVYISITTAARVSTSTIAYRPD 313  
Db 208 FKXPMLSNDNDPPDNSTVLLTLPMYTTAATNLNLHQQY-IQFAERWKSVMY-----D 261  
QY 314 SSFM--KSIMATQLRDLATWYTTLYRQNPFCPSRNRRTAVSBFMKXNTHVLIRNET-- 368  
Db 262 ESFINQTKVDLQRIQDYSTTVSTTFE-KFKPTLNP-SNKESVKNYR-----YVRSMTLQ 315  
QY 369 -----PYTYGTLDMSLYNETMFV-----ENKTASDNKTTPTSPSMG 408  
Db 316 SLDAIAATWPTLDNVNPSNVDIQLDQTRLVFSDVAGPWEGNDNITSNIIDVLTPTINTGIG 375  
QY 409 FQ-----RTFIDPLMDYLDLSLLFLDEIRNFSRLSPYVNLTPPEH 448  
Db 376 FOESSDLRKFTYPRIE-LQSMQPHGQ-----YVNSKSVEH 409  
RESULT 3  
US-08-485-568A-6  
; Sequence 6, Application US/08485568A  
; Patent No. 5589382  
; GENERAL INFORMATION:  
; APPLICANT: Payne, Jewel  
; APPLICANT: Narva, Kenneth E.  
; APPLICANT: Fu, Jenny M.  
; TITLE OF INVENTION: No. 5589382el Bacillus thuringiensis Genes Encoding  
; TITLE OF INVENTION: Nematode-Active Toxins  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,568A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/310,197  
; FILING DATE: 21-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/918,345  
; FILING DATE: 21-JUL-1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/558,738  
; FILING DATE: 27-JUL-1990  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/357,698  
; FILING DATE: 16-DEC-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/176,403  
; FILING DATE: 30-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/999,053  
; FILING DATE: 14-JUL-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.

```
;
;
;   REGISTRATION NUMBER: 31,794
;   REFERENCE/DOCKET NUMBER: MA48DD2.C1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
;   INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1167 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-485-568A-6

Query Match      4.3%; Score 106.5; DB 1; Length 1167;
Best Local Similarity 20.9%; Pred. No. 0.28;
Matches 97; Conservative 67; Mismatches 132; Indels 169; Gaps 26;

Qy 90 PPQY-ILAGPIQNSITYLWDFYSTQLRKPQKYYVSYQVNHAKTITFRPPCGRVPSM 148
Db 8 YPSYHNVLAHPRLDS-----FFDPF-VETFKDLGAWEEF-----GKTGYM 48
Qy 149 TCLSEMLNVSKRNDTGEQCGNFTTFNPMFNVPRMNTKLVGPTKVNVD-----SQ 200
Db 49 DPLKQHLQIAW--DTSQNG-----TVDYLAUTKASISLIGLIPGADA 88
Qy 201 TIYPLGL-----TALLRYAQRNCTHSFY-----LVNMSRN--LFRVPKYING- 242
Db 89 VVPIINMFVDFIPKLFGRGSGQNAQAQFFELIIEKVELVDEDFRFTLNLLNYLDGM 148
Qy 243 -----TKLQNTWR-----KLKQKQAP-----VKEQPEKAKKTQSTTTPY 277
Db 149 QTALSHFQNDVQIAICQEQGLMDQTPTACTPTDHLISVRESF-KDARTTITETALPH 207
Qy 278 FS-----YTTSAALNVTTNVTYSITTAARRVSTSTIAVRPD 313
Db 208 FKNPMLSTNDNTPDFNSDTVLLTLPMYTTGATNLILHOGY-IQFAERWKSVMY-----D 261
Qy 314 SSFW---KSMATQLRLATVYTTLYRQNPFCPSRRNRNAVSEFMKNTHVLRNET-- 368
Db 262 ESFINQTKVLDORRIQDYSTVSTTFE-KFKPTLNPS-NKESVVKYNR-----YVRSMTLQ 315
Qy 369 -----PYTYIGTLDMSLYNMTFV-----ENKTASDSNKTTPSPSMG 408
Db 316 SLDIAATWPTLDNVNPSNDVIQLDQTRLVPSDVAGPWEQGNIDNITSNIIDVLTPINTGIG 375
Qy 409 FQ-----RTIDPLWDYLDLSLLFDEIRNFSRLSPTVYVNLTPPEH 448
Db 376 FOESDLRKFTYPRIE-LQSMQFHGQ-----YVNSKSVEH 409

RESULT 4
US-08-590-554A-6
; Sequence 6, Application US/08590554A
; Patent No. 5831011
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Fu, Jenny M.
; TITLE OF INVENTION: No. 5831011el Bacillus thuringiensis Genes Encoding
; TITLE OF INVENTION: Nematode-Active Toxins.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```



97	SVPTGSGCYLGVEVTPPVGSGTTQN-----YGMVSSVLQANRSVISTFRWBPQSVFISDLFPY	151
114	STQLRKPAAVYYSOVNHTAKTITFRPPPCGRVPMSMTCISEMLNVSKRNDTGCGGCGNP--	171
152	-PEIRGVGVAIYNKSNPKGAFIVLAPSGYG-----LLRYKRANDTVENASFSFSR	200
172	--TFENPMFPNVPBWNTKL-----YV-GPTKYN-----VDSOTIYFL	205
201	SKTTAMTNMYNFPNSNDELGSGLLPRCGSNFBSARYVDSGDKVHSITLTISSYSAVI	260
206	G-LPALLLLRYAQRNCTHSFYLVNMSRNLFRVPKYINGTKLTKNTKLR-----KQAPVK	260
261	GPLEA--GEVAMRD--DSKFYAVIRTSGGSFARSKTLSSMDNVNIKALKACDVIPTSTN	317
261	EQPEKKAKTKTOSTTT-----PYPS--YTTTSAALN-----VTTNVT	293
318	IQPPTQAKNYATPTKLADNLASIVSNCPYAKNIYLTLSFNNLVSGSETGMELTST	377
294	YSINT-----AARRVSTSTIAIRPDSSPFMKSIM	321
378	ENITLTPVWVSLKSAOSNICOANAQDALRVLGSKLACQNNKSTONIL	427

RESULT 7  
US-09-248-796A-23675  
; Sequence 23675, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 23675  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-23675

RESULT 8  
US-09-248-796A-18994  
; Sequence 18994, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANT

```

, , TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
, , FILE REFERENCE: 107196.132
, , CURRENT APPLICATION NUMBER: US/09/248,796A
, , CURRENT FILING DATE: 1999-02-12
, , PRIOR APPLICATION NUMBER: US 60/074,725
, , PRIOR FILING DATE: 1998-02-13
, , PRIOR APPLICATION NUMBER: US 60/096,409
, , PRIOR FILING DATE: 1998-08-13
, , NUMBER OF SEQ ID NOS: 28208
, , SEQ ID NO 18994
, , LENGTH: 476
, , TYPE: PR1
, , ORGANISM: Candida albicans
, , FEATURE:
, , NAME/KEY: UNSURE
, , LOCATION: (345)
, , OTHER INFORMATION: Identity of amino acid sequences
US-09-248-796A-18994

```

RESULT 9  
US-08-415-593-43  
; Sequence 43, Application US/08415593  
; Patent No. 5912140  
; Patent No. 5912140 5776726  
; GENERAL INFORMATION:  
; APPLICANT: Whoriskey, Susan K.  
; APPLICANT: Quinn, Cheryl L.  
; APPLICANT: Tao, Niajun  
; APPLICANT: Politis-Virk, Karen I.  
; APPLICANT: Schimmel, Paul R.  
; TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA  
; TITLE OF INVENTION: Synthetase Genes, Tester Strains and Assays  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.  
; ZIP: 02173



ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/097,889  
FILING DATE: 15-JUN-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rosenman Ph.D., Stephen J.  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 660088.417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 347 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-097-889-14

Query Match 3.9%; Score 96.5; DB 3; Length 347;  
Best Local Similarity 22.3%; Pred. No. 0.43;  
Matches 68; Conservative 33; Mismatches 109; Indels 95; Gaps 12;

Qy 170 NFFTPNPFNNVPRNTKLYVGPVKVNDVSDTIYFLGJTALL-----RYAQRNCTH-- 221  
Db 36 NMLAFIPVL--TKQNPSTEARAIKYFTQATASMLLMALFNMLSGQMTMTNTTQY 93

Qy 222 -SFVLVNAISRNL-----FRVPKYINGTKLNTWKRLK-RKQAPVKEQFEKAKKTOST 273  
Db 94 SSLMIMMAMAKMGAPFHFVPEVTQGTPLTSLGLLLTWKLPAPIS----- 140

Qy 274 TTPFSTYTSALNVTNTVYSITTAARRVSTSTIAYRPSDSSFMKSIMA-----TOL 325  
Db 141 ----IMYQISPLSNVSLLLTSLIL-----SIMAGSWGGLNQTL 175

Qy 326 RDLAT-----WVYTLRYQRNPFCEPSRNRRTAVSEFMKNTVLIENETPYTYIGTLD 377  
Db 176 RKILAYSSITMGMMMAVLPPY-----NPNNTILNLTIIYILITTA 215

Qy 378 MSSLYYNE--TMFVENKTASDNKTTPTSPSMGFORTFIDPLWDYLDLSLLFLDEI-RNFS 434  
Db 216 FLLNLNSSTTLLLSRTWNLKLTWLTPLIPSTLLSLGGLPLTGLPKWAIIEFTKNS 275

Qy 435 LRSPT 439  
Db 276 LIIPT 280

RESULT 12  
US-09-098-079-14  
Sequence 14, Application US/09098079  
Patent No. 6489095  
GENERAL INFORMATION:  
APPLICANT: Hernstadt, Corrina  
APPLICANT: Ghosh, Soumitra S.  
APPLICANT: Clevenger, William  
APPLICANT: Fahy, Eoin P.  
APPLICANT: Davis, Robert E.  
TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF  
TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/098,079  
FILING DATE: 15-JUN-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rosenman Ph.D., Stephen J.  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 660088.416  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 347 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-098-079-14

Query Match 3.9%; Score 96.5; DB 4; Length 347;  
Best Local Similarity 22.3%; Pred. No. 0.43;  
Matches 68; Conservative 33; Mismatches 109; Indels 95; Gaps 12;

Qy 170 NFFTPNPFNNVPRNTKLYVGPVKVNDVSDTIYFLGJTALL-----RYAQRNCTH-- 221  
Db 36 NMLAFIPVL--TKQNPSTEARAIKYFTQATASMLLMALFNMLSGQMTMTNTTQY 93

Qy 222 -SFVLVNAISRNL-----FRVPKYINGTKLNTWKRLK-RKQAPVKEQFEKAKKTOST 273  
Db 94 SSLMIMMAMAKMGAPFHFVPEVTQGTPLTSLGLLLTWKLPAPIS----- 140

Qy 274 TTPFSTYTSALNVTNTVYSITTAARRVSTSTIAYRPSDSSFMKSIMA-----TOL 325  
Db 141 ----IMYQISPLSNVSLLLTSLIL-----SIMAGSWGGLNQTL 175

Qy 326 RDLAT-----WVYTLRYQRNPFCEPSRNRRTAVSEFMKNTVLIENETPYTYIGTLD 377  
Db 176 RKILAYSSITMGMMMAVLPPY-----NPNNTILNLTIIYILITTA 215

Qy 378 MSSLYYNE--TMFVENKTASDNKTTPTSPSMGFORTFIDPLWDYLDLSLLFLDEI-RNFS 434  
Db 216 FLLNLNSSTTLLLSRTWNLKLTWLTPLIPSTLLSLGGLPLTGLPKWAIIEFTKNS 275

Qy 435 LRSPT 439  
Db 276 LIIPT 280

RESULT 13  
US-09-107-532A-4902  
Sequence 4902, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts

```
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4902:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 651 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...651
; SEQUENCE DESCRIPTION: SEQ ID NO: 4902:
US-09-107-532A-4902

Query Match          3.9%; Score 95.5; DB 4; Length 651;
Best Local Similarity 20.7%; Pred. No. 1.5;
Matches 87; Conservative 34; Mismatches 121; Indels 179; Gaps 22;

Qy 71 STKYNVSKQPVKNLTMNTEP-----PQYVILAGPIQ-NYSITLWFDVYSTQLR-----K 119
Db 43 ATRYN-----KMLFLPTSPKNGSEY-----PIQYNYCINPFCNFGKQHKFKDVKGK 91
Qy 120 PAKY--VYSQYNHTAKTITFRPPPCGRVPSMTC-----LSEMLNVSKRNDT- 163
Db 92 PSRYKMTGSKDKGHKGMYCNDNPTIGRGVSONCTVTPLSNWSVVVEIKRLIEINSIQDVE 151
Qy 164 -----GEOGCG--NFTTFN----- 175
Db 152 PDYQFHKEGCEESTPNEPKQFYKRGSKRGKSQRYOCKACKFTNVLPKRETTTHQ 211
Qy 176 -----PMFNVPRWNTKLYGPTKVNVDSTIYFLGLTA-----LLRYAQRNC----- 219
Db 212 QKNTILPMF-----AKWVGRVSV-----SRTCDILGIGVGYTHKLEWLYRRCLEFLER 261
Qy 220 -----THSFVLVNAAGRNLFRVPKYINGTKLQNTWRKLRKQAPVKQFEKKAKKTQS 272
Db 262 YETQPLQTKKF--NEMWLNTDKMYYLNNVRKKG-----QGSKK----- 298
Qy 273 TTTPVFSYTTSAALNVTNTVYSITTAARRVSTSTIAY-----RPDSSFMSKSI 322
Db 299 -----YTGFDLNMQTYIVVSAEVLRYVFRSDVAYDWNISMDELNEDTRKFKEDHLN 351
Qy 323 TQLR--DLATWVYTTLRVQRNPFCEPSNRRTAV-----SEFMKNTHLVLRNETPY 370
Db 352 TFSRKNDRLDWSY-----YQEPSANDSENAYLHELKGIKTNRSRFRVDGLNV----DAPY 403
Qy 371 T 371
Db 404 T 404
```

```
RESULT 14
US-09-134-000C-5699
; Sequence 5699, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5699
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5699

Query Match          3.9%; Score 95.5; DB 4; Length 651;
Best Local Similarity 20.7%; Pred. No. 1.5;
Matches 87; Conservative 34; Mismatches 121; Indels 179; Gaps 22;

Qy 71 STKYNVSKQPVKNLTMNTEP-----PQYVILAGPIQ-NYSITLWFDVYSTQLR-----K 119
Db 43 ATRYN-----KMLFLPTSPKNGSEY-----PIQYNYCINPFCNFGKQHKFKDVKGK 91
Qy 120 PAKY--VYSQYNHTAKTITFRPPPCGRVPSMTC-----LSEMLNVSKRNDT- 163
Db 92 PSRYKMTGSKDKGHKGMYCNDNPTIGRGVSONCTVTPLSNWSVVVEIKRLIEINSIQDVE 151
Qy 164 -----GEOGCG--NFTTFN----- 175
Db 152 PDYQFHKEGCEESTPNEPKQFYKRGSKRGKSQRYOCKACKFTNVLPKRETTTHQ 211
Qy 176 -----PMFNVPRWNTKLYGPTKVNVDSTIYFLGLTA-----LLRYAQRNC----- 219
Db 212 QKNTILPMF-----AKWVGRVSV-----SRTCDILGIGVGYTHKLEWLYRRCLEFLER 261
Qy 220 -----THSFVLVNAAGRNLFRVPKYINGTKLQNTWRKLRKQAPVKQFEKKAKKTQS 272
Db 262 YETQPLQTKKF--NEMWLNTDKMYYLNNVRKKG-----QGSKK----- 298
Qy 273 TTTPVFSYTTSAALNVTNTVYSITTAARRVSTSTIAY-----RPDSSFMSKSI 322
Db 299 -----YTGFDLNMQTYIVVSAEVLRYVFRSDVAYDWNISMDELNEDTRKFKEDHLN 351
Qy 323 TQLR--DLATWVYTTLRVQRNPFCEPSNRRTAV-----SEFMKNTHLVLRNETPY 370
Db 352 TFSRKNDRLDWSY-----YQEPSANDSENAYLHELKGIKTNRSRFRVDGLNV----DAPY 403
Qy 371 T 371
Db 404 T 404

RESULT 15
US-09-538-092-84
; Sequence 84, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:21:31 ; Search time 164 Seconds  
(without alignments)  
1017.129 Million cell updates/sec

Title: US-09-942-146A-1  
Perfect score: 2451  
Sequence: 1 MGRKEMVRDVPFMFVLISI.....PHRRAVNLTSGNSLWMLQ 465  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.0  
Searched: 2002273 seqs, 358729299 residues  
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003s:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2451	100.0	465	8	ADO15374 CMV glyco
2	2420	98.7	465	7	ADB81330 Human cyt
3	710	29.0	31267	6	ABG74786 Human RGS
4	681	27.8	26926	8	ADQ17316 Human sof
5	678	27.7	26926	4	AAU05396 Human tit
6	581	23.7	16368	6	ABM67171 Photorhab
7	580	23.7	11300	5	ABB84277 Adenoviru
8	575	23.5	15281	2	AAR44929 T. niveum
9	561	22.9	22157	8	ADP84155 Human CA1
10	539	22.0	8601	6	AAO29660 Paederus
11	538	22.0	19938	6	ABP76678 Streptomy
12	536	21.9	19938	6	ABP76681 Streptomy
13	534	21.8	19938	6	ABP76679 Streptomy
14	531	21.7	19938	6	ABP76680 Streptomy
15	526	21.5	7107	4	ABBS58144 Drosophil
16	522	21.3	19938	6	ABBS98398 Streptomy
17	518	21.1	6940	8	ADN96828 Bugula br
18	516	21.1	10182	5	ABP38334 Staphyloc
19	514	21.0	10203	8	ADO84851 S. epiderm
20	508	20.7	19938	6	ABP76682 Streptomy
21	507	20.7	7052	6	AAE35266 Human P45
22	499	20.4	11721	6	ABUS4721 Human CA1
23	486	20.2	7718	4	ABG11811 Novel hum
24	492	20.1	6619	4	ABG23329 Novel hum
25	490	20.0	7036	6	AAE35264 Human P45

## ALIGNMENTS

RESULT 1  
ADO15374  
ID ADO15374 standard; protein; 465 AA.

XX ADO15374;

XX 01-JUL-2004 (first entry)

XX CMV glycoprotein O (go) polypeptide.

XX CMV; glycoprotein O; go; glycoprotein H ; glycoprotein L; glycoprotein B;  
XX CMV infection; tumour; cancer; virucide.

XX Cytomegalovirus.

XX US2004013682-A1.

XX 22-JAN-2004.

XX 26-JUN-2002; 2002US-00942146.

XX 29-JUL-1999; 99US-0146180P.

XX 28-JUL-2000; 2000US-00627986.

XX (COMP/) COMPTON T.

XX (HUBE/) HUBER M T.

XX Compton T, Huber MT;

XX WPI; 2004-121532/12.

XX Designing new anti-CMV drug by analyzing binding of glycoprotein O to glycoprotein O receptor, designing candidate drug that would competitively interfere with glycoprotein O binding to glycoprotein O receptor.

XX Disclosure; Fig 3; 7pp; English.

XX The invention relates to a method of designing a new anti-cytomegalovirus (CMV) drug, involving analysing the binding of glycoprotein O to a glycoprotein O receptor, designing a candidate drug that would competitively interfere with glycoprotein O binding to the glycoprotein O receptor and showing that the candidate drug competitively inhibits glycoprotein O binding to the glycoprotein O receptor. The invention also relates to a method of screening involving determining whether a candidate drug interferes with a glycoprotein O-containing complex binding to a cell surface. The evaluation involves analysing the inhibition of major early protein of the human cytomegalovirus (HCMV).

Aae35265 Human P45  
Abg76187 Human ser  
Abg21064 Novel hum  
Adl91934 Streptomy  
Aay77180 S. venezu  
Abg23295 Novel hum  
Abu54708 Human CA1  
Abu33707 Protein e  
Aao29662 Paederus  
Abb67112 Drosophil  
Aae36132 Streptomy  
Aau85008 Human mel  
Aau54861 Human CA1  
Aau84596 HIV compl  
Aag94939 Shrimp wh  
Adg70546 Aspergill  
Abu48941 Protein e  
Abu48947 Protein e  
Abg08561 Novel hum  
Abj19119 Pathogen

26 489 20.0 7067 6 AAE35265  
27 488 19.9 7368 6 ABG76187  
28 483 19.7 9222 4 ABG21064  
29 483 19.7 11877 8 ADL91934  
30 483 19.7 12199 3 AAY77180  
31 480 19.6 5909 4 ABG23295  
32 476 19.4 9799 6 ABUS4708  
33 475 19.4 10421 6 ABU33707  
34 473 19.3 6266 6 AAO29662  
35 473 19.3 8805 4 ABB67112  
36 472 19.3 10917 6 AAE36132  
37 469 19.1 5546 5 AAU85008  
38 468 19.1 10431 6 ABUS4861  
39 468 19.1 5746 5 AAU84596  
40 468 19.1 6025 4 AAG94939  
41 466 19.0 7064 7 ADG70546  
42 465 19.0 4688 6 ABU48941  
43 465 19.0 5005 6 ABU48947  
44 465 19.0 5266 4 ABG08561  
45 465 19.0 10498 6 ABJ19119

CC The invention also relates to a vaccine useful for diminishing CMV  
CC infection and comprising at least a fragment of the glycoprotein O  
CC polypeptide in combination with a carrier. The vaccine comprises at least  
CC a fragment of a glycoprotein chosen from CMV glycoprotein H, CMV  
CC glycoprotein L, or CMV glycoprotein B, and is useful for vaccinating a  
CC patient against CMV infection. The drug and the vaccine are useful for  
CC treating CMV infections in tumours. This sequence represents the CMV  
CC glycoprotein O (gO) polypeptide of the invention.  
XX  
SQ Sequence 465 AA;

Query Match 100.0%; Score 2451; DB 8; Length 465;  
Best Local Similarity 100.0%; Pred. No. 2.6e-63;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MGRKEMVRDVPKMFVLISISFLVSVFNCVKMSKALYNRPWRGLVLSKIGKYLDDQL 60  
DB 1 MGRKEMVRDVPKMFVLISISFLVSVFNCVKMSKALYNRPWRGLVLSKIGKYLDDQL 60  
QY 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEPPQYVILAGPIQNYSTIYLFDFYSTQLRKP 120  
DB 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEPPQYVILAGPIQNYSTIYLFDFYSTQLRKP 120  
QY 121 AKYVYSQVNHHTAKTITPRPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTFNPMFFN 180  
DB 121 AKYVYSQVNHHTAKTITPRPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTFNPMFFN 180  
QY 181 VPRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSFVLNVAMSRNLFPRVPKYI 240  
DB 181 VPRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSFVLNVAMSRNLFPRVPKYI 240  
QY 241 NGTKLNTMRKLKQAPVKEQEKAKKTQSTTTPYFSTTSAAALNVNTNVTYSITTA 300  
DB 241 NGTKLNTMRKLKQAPVKEQEKAKKTQSTTTPYFSTTSAAALNVNTNVTYSITTA 300  
QY 301 RRVSTSTIAYRPDSSFMKSIATQLRDLATVWVYTLRYQNPCEPSRNRATVSEPMKN 360  
DB 301 RRVSTSTIAYRPDSSFMKSIATQLRDLATVWVYTLRYQNPCEPSRNRATVSEPMKN 360  
QY 361 HVLRNETPTVIYGLDMSSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWDY 420  
DB 361 HVLRNETPTVIYGLDMSSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWDY 420  
QY 421 LDSLLFLDEIRNFSLSRSPTYVNLTPPEHRRVAVNLSTNSLWMLQ 465  
DB 421 LDSLLFLDEIRNFSLSRSPTYVNLTPPEHRRVAVNLSTNSLWMLQ 465

RESULT 2  
ADB81330  
ID ADB81330 standard; protein; 465 AA.  
XX AC ADB81330;  
XX DT 04-DEC-2003 (first entry)  
XX XX Human cytomegalovirus glycoprotein O polypeptide.  
XX XX glycoprotein O; gO; anti-cytomegalovirus; CMV; AIDS;  
XX KW bone marrow transplant; glycoprotein H; glycoprotein L; gene therapy.  
XX OS Human herpesvirus 5.  
XX XX US6569616-B1.  
XX PN 27-MAY-2003.  
XX PD 28-JUL-2000; 2000US-00627986.  
XX PF 29-JUL-1999; 99US-0146180P.  
XX PR (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX PA  
XX XX

PI Compton T, Huber MT;  
XX WPI; 2003-605462/57.  
DR  
XX Screening candidate anti-cytomegalovirus (CMV) drugs for the ability to  
PT block CMV entry into a host cell comprises determining whether a  
PT candidate drug interferes with glycoprotein O binding to glycoprotein O  
PT receptor.  
XX  
PS Disclosure; Fig 3; 8pp; English.  
XX  
CC This invention relates to a novel method of designing a new anti-  
CC cytomegalovirus (CMV) drug that has the ability to block CMV entry into a  
CC host cell. CMV is a member of the Herpesviridae family and is an  
CC opportunistic pathogen responsible for serious clinical disorders in  
CC immunosuppressed patients, such as persons with AIDS or those with recent  
CC organ or bone marrow transplants. The present invention describes a new  
CC drug target that comprises a CMV viral component designated glycoprotein  
CC O (gO), whereby the candidate drug works by interfering with gO binding  
CC to its receptor, where glycoprotein O is part of a complex with  
CC glycoprotein H and glycoprotein L. Accordingly, using gene therapy or  
CC developing antibodies to go that act as vaccines, it is possible to block  
CC virus infection by impeding the viral entry pathway. Furthermore, the  
CC method is useful in identifying alternative drug targets and immunogens  
CC that elicit protective activity against CMV infection. This polypeptide  
CC is the human cytomegalovirus glycoprotein O amino acid sequence of the  
CC invention.  
XX  
SQ Sequence 465 AA;

Query Match 98.7%; Score 2420; DB 7; Length 465;  
Best Local Similarity 99.4%; Pred. No. 2.2e-62;  
Matches 463; Conservative 0; Mismatches 1; Indels 2; Gaps 2;  
QY 1 MGRKEMVRDVPKMFVLISISFLVSVFNCVKMSKALYNRPWRGLVLSKIGKYLDDQL 60  
DB 1 MGRKEMVRDVPKMFVLISISFLVSVFNCVKMSKALYNRPWRGLVLSKIGKYLDDQL 60  
QY 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEPPQYVILAGPIQNYSTIYLFDFYSTQLRKP 120  
DB 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEPPQYVILAGPIQNYSTIYLFDFYSTQLRKP 120  
QY 121 AKYVYSQVNHHTAKTITPRPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTFNPMFF 179  
DB 121 AKYVYSQVNHHTAKTITPRPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTFNPMFF 180  
QY 180 NVRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSFVLNVAMSRNLFPRVPKY 239  
DB 181 NVRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSFVLNVAMSRNLFPRVPKY 240  
QY 240 INGTKLNTMRKLKQAPVKEQEKAKKTQSTTTPYFSTTSAAALNVNTNVTYSITTA 299  
DB 241 INGTKLNTMRKLKQAPVKEQEKAKKTQSTTTPYFSTTSAAALNVNTNVTYSITTA 300  
QY 300 ARRVSTSTIAYRPDSSFMKSIATQLRDLATVWVYTLRYQNPCEPSRNRATVSEPMKN 359  
DB 301 ARRVSTSTIAYRPDSSFMKSIATQLRDLATVWVYTLRY-QNPCEPSTNRATVSEPMKN 359  
QY 360 THVLIRNETPTVIYGLDMSSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWD 419  
DB 360 THVLIRNETPTVIYGLDMSSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWD 419  
QY 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRRVAVNLSTNSLWMLQ 465  
DB 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRRVAVNLSTNSLWMLQ 465

RESULT 3  
ABG74786  
ID ABG74786 standard; protein; 31267 AA.  
XX AC ABG74786;  
XX XX

DT	05-JUN-2003	(first entry)	12	-----PK-----	13
XX	DE	Human RGS11 protein.	629	SLTEQAGPEPAAPYFITKPVVQKLVEGGSVVFGQVGNPKPHVYWKSGVPLTTGYR	688
XX	KW	RGS11; human; screening; cardiant; antiangiinal; gene therapy;	14	-----	13
XX	KW	heart disorder; cardiac ischaemia; heart failure; angina.	689	YKVSYNKQTGECKLVISMTFADDAEYTIIVRNKHGETSASASLLEADYELLKMSQEM	748
OS	XX	Homo sapiens.	14	-----	13
XX	PN	WO2002103355-A1.	749	LYQTQVTAFOBPVEGETAPGFVYSEYEKEQALIRKIMAKDTVVVRYVEDQEFHI	808
XX	PD	27-DEC-2002.	14	-----	13
XX	PF	17-JUN-2002; 2002WO-JP006019.	809	SSFEERLIKEIYRIIKTTLEELLEEDGEEKMAVDISESEAVESGFDLRIKNRYILEGMG	868
XX	PR	18-JUN-2001; 2001JP-00183038.	14	-----	13
XX	PA	(TAKE-) TAKEDA CHEM IND LTD.	869	CHFSLDQVWISITKDOKRIKHGERYQMDFLQDGRASLRIPVVLPEDEGIYTAFASNIKN	928
XX	PI	Koyama N, Tanida S, Yamamoto K;	14	-----	13
XX	DR	WPI; 2003-167557/16.	929	AICSGKLYVEPAAPLGAFTYIPTLEBPVSIRISLSPRSVSRSPIRMSPARMSPARM	988
XX	DR	N-PSDB; ABX13540.	14	-----MFVLISISF-----	22
XX	PT	Screening compounds regulating RGS11 expression and activity for	989	SPARMSPGRRLEBETDESQLERLYKPVFLKPVSKFCLEGQTARFDLKVGRMPETFWPH	1048
XX	PT	prevention and treatment of heart disease.	23	-----	22
XX	PS	Claim 1; Page 59-261; 321pp; Japanese.	1049	DQOQIWNVTHKVIKEDGTQSLIIVPATPDSGEVTVVQNRAGRSSISVILTVEAVEH	1108
XX	CC	This invention describes a novel method for screening compounds for their	23	-----	22
XX	CC	ability to regulate the activity and expression of human RGS11 and its	1109	QVKPMPEVKLVKNVNIKEGSRLEMKVATGNPNPDIVLWLNDSIIIVPHKYPKIRIEGTKE	1168
XX	CC	partial peptides and salts, by observing the expression or activity of	23	-----LLV-----	25
XX	CC	RGS11 in the presence or absence of the test compound. The products of	1169	AALKIDSTVSQDSAWYTATATAINKAGRDTTRCKVNVVEFAEPERPKLIIPRGTYRAKEI	1228
XX	CC	the invention have cardiant and antiangiinal activity and can be used for	26	-----	25
XX	CC	gene therapy. The methods and compositions are useful in the prevention,	1229	AAPELEPLHLRYGQEQWEGDLYDKEKQKPFKKLTSRLRKRFQPAHPECLTPIGDP	1288
XX	CC	treatment and diagnosis of heart disorders such as cardiac ischaemia,	26	-----SFINCKVMSK-----	35
XX	CC	heart failure and angina. This sequence represents the human RGS11	1289	TWVVEWLHDGKPLEAANRLRMINEFGYCSLDYGVAYSRSDSGIITCRATNKYGTDTSATL	1348
XX	CC	protein described in the disclosure of the invention	36	-----ALYNRPWR-43	43
SQ	Sequence 31267 AA;		1349	IVKDEKSLVEESQLPEGRKGLORIEBELERMAHEGALTGVTTDOKEKQKPDIVLYPEPVRV	1408
			44	-----	43
			1409	LEGETARFCRVGTGYPQPKVNMVNLGQLIRKSRFRVRYDGIHVIDVCKSYDTGEVKV	1468
			44	-----GLVLSKI-----	50
			1469	TAENPEGVIEHKVLEIKQREDFRSVLRAPRPRPEFHVHEPKLQFEYQKVDPRPVDTE	1528
			51	-----	50
			1529	TKEVVKLKAERITHEKVPSESEELRSKFRRTTEGYEAITAVELKSRKODESEBELLR	1588
			51	-----	50
			1589	KTKDELLHWTKLTBEEKKALAEKGKITTPTPKPKIELSPSMEAPKIFERIQSQTVGQ	1648
			51	-----	50
			1649	SDAHFRVVRVKGPDCEWEYKNGVKIERSDRIYVWYWPEDNVCELVIIRDVTAEDSASIMVK	1708
			51	-----	50

Query Match 29.0%; Score 710; DB 6; Length 31267;  
 Best Local Similarity 0.9%; Pred. No. 2.5e-07;  
 Matches 271; Conservative 92; Mismatches 98; Indels 30346; Gaps 87;

QY	3	RKE-----	5
DB	209	RKEAKTALSTIAVATAKAKEQETILRTRETMTATROEQIQTHTGKVDVGGKAEAVTVVA	268
QY	6	-----	5
DB	269	AVDQARVREPREPHLEESVAQQTILEYGYKERISAAKVAEPPQRPASEPHVVPKAVKPR	328
QY	6	-----	5
DB	329	VIOAPSETHIKTTDQGMHISSQIKKTTDLTTERLVHVDKRPRTASPHFTVSKISVPKTE	388
QY	6	-----	5
DB	389	HGYEASIGASAIATLOKELSATSQAOKIKYSVKAPTUKVSETVRABEPTLPQFPADTP	448
QY	6	-----	5
DB	449	DTYKSEAGVEVKVEGVSITGTTVREERPEVLHGREAKVTETARVPAPVEIPTPTPLVS	508
QY	6	-----MMVRDV--11	11
DB	509	GLKNVTIEGESVLECHISGSPSTVWYREDYQIESIDFQITQSGIARLMIREAFA	568
QY	12	-----	11
DB	569	EDSGRTCSAVNAGTVSTSCYLAQVSEFEKETTAETKFTTBKRFVESRDVMTDT	628

Db	1709	AINIAGETSSHAFLVQAKLITFQELQDVVAXEKOTMATFECETSEPFVKVWKYKDG	1768	Db	2789	QPARFCRVSGTDLKVSWSKDKKIKPSRFRMTQFEDTVQLEIAEAYPEDEGTYTFV	2848
Qy	51	-----	50	Qy	102	-----	101
Db	1769	EVHGDKYRMHSDRKVHFLSILTIDTSDAEDYSCVLVEDENVKTTAKLIVEGAVVEFVKE	1828	Db	2849	NAVQVSSSTANLSLEVSFKKEPLGQKPSFIQPLSSLRVNHGETVRFHARVSGIPKPEIQW	2908
Qy	51	-----GKY-----KLD	56	Qy	102	-----	101
Db	1829	LODIEVPESYSELECIVSPENIBGKWHNDVELKNGKYITISRGQRNLTVKDVTKED	1888	Db	2909	FHNQQLILPTKDVVHFPESTGMALMLIIVDAVSEHAGQYSCKAANSAGEATCAATLTVTP	2968
Qy	57	Q-----	57	Qy	102	-----	101
Db	1889	QGEYSFVIDGKTKTKLMMKPRPAILQGLSDQKVCBGDIVQLEVKVLSLESGVGMKOG	1948	Db	2969	KKAASVDQTQCHTASLSQIAESTELSKAKESTDDSFIDVTWTHGAKIEBERLKQSQ	3028
Qy	58	-----LKLLELRQL-	66	Qy	102	-----	101
Db	1949	QEVQPSDRVHIVIDKQSHMLLIJEDMTKEDAGNYSFTIPALGLSTSGRVSVYKIKILRGLR	2008	Db	3029	NGNIQPLTTCNVQLVDQGLYSCIVHNDGCBRTTSAVLSVEDVNVNTOARGOLEEIMANTT	3088
Qy	67	-----ETTISTKYNVS-----	77	Qy	102	-----	101
Db	2009	DLTCTQNVVFEVELSHGIDVLWNFKDKEIKPSSKYKIEAHGKIYKLTVLNMMKDEG	2068	Db	3089	IGESSTYITNVCEKEBVYETPENVSQAITPHASESGTLVNVVESEIEIASERIKKDDLRE	3148
Qy	78	-----	77	Qy	102	-----	101
Db	2069	KYTFYAGENMTSGKLTVAGGAISKPLTDQTVASSQEAVFCEVANPDSKGWLRDGHLP	2128	Db	3149	LQLSASTRIDEPKTEQKEENMRFENSRFRQRCPPSPLOEIESQEVYEGDCNPFVCHF	3208
Qy	78	-----KQPVKNLTM-----	86	Qy	102	-----NYSITVL-----	108
Db	2129	LTNNIRTVKIKTKLNLTVTQDAVFTVELTHPNVKGQVQWIKNGVVLESNEKYAISVG	2188	Db	3209	QGYQPQIVTWYNDMPPIPRNQNFIIHLSLENYISILTSSVHHQNEGSITCVLPNQYGTVKT	3268
Qy	87	-----	86	Qy	109	-----	108
Db	2189	TIYSLRIKCAIVDESIVYFRLGRLGASARLHVETVKIIKKPKDVTALENATVAFEVSVS	2248	Db	3269	TSMLKVKAKQKHDVKAHKVPVPHDYLDDEEELAVFDQAKGAHPMSQEGQTNLHLLKTN	3328
Qy	87	-----	86	Qy	109	-----	108
Db	2249	HDTVVKWFKHSVEIKPSDKHRLVSRKVKHKLMLQNISPSDAGEYTAVVGQLECKAKLFV	2308	Db	3329	PPVPPSGDTCELLSPFVEIQVTAATPIPBQDKESKEVFQTELEPKAMPQDQVTPKGRF	3388
Qy	87	-----NTEFPQ-----	92	Qy	109	-----WF-----	110
Db	2309	ETLHITKMNIEVPETKTASPECEVSHFNVPMSWLKNGVEIEMSEKFKIVVQGLHQLI	2368	Db	3389	VPLSDITNEPPKMLQEMPKHAKCREGDSIIILECLISGEPQPVVTVWFQNGVLLKQKQKF	3448
Qy	93	-----	92	Qy	111	-----	110
Db	2369	IMNTSTDSABYTFVCGNDQVSATLTVTPIMITSMMLKDINAEBKDTITFEVTVNYEGISY	2428	Db	3449	EEVNGSHQLYIKDVNSQDSGKYKCAENNSGAVESVSDLTVEPVTYRENSQENIGEIVG	3508
Qy	93	-----	92	Qy	111	-----	110
Db	2429	KWLKNGVEIKSTDKCQWRTKCLTHSLIIRNVHFGDAADYTFVAGKATSTATLYVEARHIE	2488	Db	3509	KYSRDQQLQDQGESVRAHFYDYPAGPFTFTWNVKEYSVRDYFQSLLETIEQIDQKEQREVP	3568
Qy	93	-----	92	Qy	111	-----	110
Db	2489	FRKHIDIKVLEKKRAMFECEVSEPDITVQMMKDDQELQITDRIKIQEKEYVHRLLIPT	2548	Db	3569	ETKDRDQAKLVQSESIITSMEEVETFTVVEYNNQKQESLGRPLSPESDISIGVSTTS	3628
Qy	93	-----YVILAG-----	98	Qy	111	-----DFYSTQL-----	117
Db	2549	RMSDAGKYTVVAGNVSTAKLFVEGRDVRIRSIKKEVQVIEKQRAVVEFEVNEDDVDHAW	2608	Db	3629	BEISELDQFTPPSSVEYFESPKSPDLYFNPSDITTKQSSIHSGGETVRYSTPLGVAER	3688
Qy	99	-----	98	Qy	118	-----	117
Db	2609	YKDGIEINFQVQERHKYVVERRIHRMFISETRQSDAGEYTFVAGNRSSVTLYVNAPEPP	2668	Db	3689	YSTPSEGEVGERYSTPPGGETLERYSTPPGGETLERYSTPPGGETLERYSTPPGGETLERYSTP	3748
Qy	99	-----	98	Qy	118	-----	117
Db	2669	QVLQELQPVTVQSGKPARFCAMISGRPOPKISWYKEEQLLSTGFKCKFLHDGQEVYLLLI	2728	Db	3749	PGETLERYSTPPGALERYSTIPTGPNPTGTFKTYPSKIEREDGTNEHFFYPTTEERGA	3808
Qy	99	-----PIQ-----	101	Qy	118	-----	117
Db	2729	EAFPEDAAYTCEAKNDYGVATTSASLSVEFVSVSPQEMPVYPPIAITPLQDVTVTSEG	2788	Db	3809	YEIWRSDSGFQPNBAIBPKONEMPPSFIEPLTKRKVYENTTLGLFIVEGLPFGVKWYR	3868
Qy	102	-----	101	Qy	118	-----	117
				Db	3869	NKSILLEPDERIKMERVGNVCSLEISNIQKGEGBYCHAVNIIGEAKSPANVDIMPQER	3928

QY 118 -----RKPA----- 121  
Db 3929 VVALPPVTHQWMEFDEHTTSRTSPQEI VLEVELSEKVKFEKQKVI VTVPEFTP 3988  
QY 122 ----- 121  
Db 3989 DHKSMIVSLDVLFPNFVDPNMDREGEDKELKIDLEVFEMPPRFIMPICDFKIPENS DAV 4048  
QY 122 -----KYVYSQ----- 127  
Db 4049 FKCSVIGIPTPEVKWYKEYMCIEPDNIKYVISEKGSHTLKIRNVCLSDSATYRCRAVNC 4108  
QY 128 ----- 127  
Db 4109 VGBAICRGFLTMGDSEIFAIAKSKVTLSSLMEELVLKSNYTDSPFQVVBGPPRFIK 4168  
QY 128 ----- 127  
Db 4169 GISDCYAPIEFSSFLSAEBGLHSAELOLSKINETLELSESPVYSTKFDSEKEGTGPI 4228  
QY 128 ----- 127  
Db 4229 FIKEVSNADISMGDVATLSVTVIGIPKPKIQWFPNGVLLTPSADYKFPVFDGDDHSLILF 4288  
QY 128 ----- 127  
Db 4289 TKLEDBEYTCMASNDYGKTCISAYLKINSKGEGHDKDTETESAVAKSLEKLGPCPPHPL 4348  
QY 128 ----- 127  
Db 4349 KELPIRCAQGLPAIFEYTVVGPAPTPTVWFKENKQLCTSVYVTI IHNPNGSGTIVNDP 4408  
QY 128 ----- 127  
Db 4409 QREDSGLYICKAENMLGESTCAAEALLVLEDTMDTTPCKAKSTPEAPEDFPOTPLKGP 4468  
QY 128 -----YNHTAK----- 133  
Db 4469 VEALDSEQEIATFVKDTILKAALITEENQOLS YEHI AKANELSQLPLGAQELQSILEQD 4528  
QY 134 ----- 133  
Db 4529 KLTPESTREPLCINGSIHFQPLKEPSPNLQIQVSOQKTSKEGILMPPEPETOAVLSDT 4588  
QY 134 ----- 133  
Db 4589 EKIFPSAMSIEQINSLTVEPLKTLAEPENGYPOSSIEPPMHSLYTSVABEVLSPKEKTV 4648  
QY 134 ----- 133  
Db 4649 SDTNREQVTLQKEAQSALILQSLSAEGHVESLQSPDVMISQVNYEPLVPSEHSCTEGG 4708  
QY 134 ----- 133  
Db 4709 KILIESANPLENAGDSAVRIEBGKSLRFPALAEKQVLLKEHSDNVVMPDQIIIESKR 4768  
QY 134 ----- 133  
Db 4769 EPVAIKKQVQVRDILLSKESLLSGIPEEORLNKIQICRALQAASVASEQPLFSEWLRN 4828  
QY 134 ----- 133  
Db 4829 IEKVEEAVNITQEPHIMCMLVTSKSVTEEVITIIIEDVDPMANLKMELDALCAII 4888  
QY 134 ----- 133  
Db 4889 YEBIDILTAGPRIQQAKTSLQEEMDSFGSQKVEPITEPEVESKYLISTERVSFNVQ 4948  
QY 134 ----- 133  
Db 4949 SRVKYLDATPVTKGAVSVVDEKQDESUKPSEKESSESCTEVBATVKIQEAGGLI 5008

QY 134 ----- 133  
Db 5009 KEDGPMIHTPLVDTVSEGDIVHLTTSITNAKENVMYFENKLVPSDEKPKCLQDQNTYTL 5068  
QY 134 ----- 133  
Db 5069 VIDKVNTEHQBYCEALNDSKGTATSAKLTVVKRAAPVIRKRIEPLVALGHLAKFTC 5128  
QY 134 ----- 133  
Db 5129 EIQSPNVRFQWPKAGREIYESDKSIRSSKYISSLEILRTQVVDCEYTCASNEYGSV 5188  
QY 134 ----- 133  
Db 5189 SCTATLTVTEAVPPTFLSRPKSLTTFVGAAKFICTVTGTPTVETIWKDGAALSPSPNW 5248  
QY 134 ----- 133  
Db 5249 RISDAENKHILELSNLTIQDRGVYSCASNKFADICQAEIIIDKPHFIKELEPVQSAI 5308  
QY 134 ----- 133  
Db 5309 NKKVHLECOVDEDRKVTVTWSKDGQKLPPGKDYKICFEDKIALEIPLAKLDSGTIVCT 5368  
QY 134 -----TITPRPPP----- 141  
Db 5369 ASNEAGSSCSATVTVREPPPSFVKVDP SYMLPGESARLHCKLGSPVIQVTFWKNKE 5428  
QY 142 ----- 141  
Db 5429 LSESNPVRMYFVNSEAILDITVKVEDSGSYSCAVNDVSGSDSCSTEIVIKEPSPFIKTL 5488  
QY 142 ----- 141  
Db 5489 EPADIVRGTNALLQCEVSGTGPEISWFKDKKQIRSSKKYRLFQKSLVCLEIFSPNSAD 5548  
QY 142 -----CG----- 143  
Db 5549 VGEYECVAVANEVKGKCGCMATHLLKEPPTFVKVDDLIALLGGQTVTLQAAVRGSEPI SVTW 5608  
QY 144 ----- 143  
Db 5609 MKQEVIREDGKIKMSFSGNVAVLIIIPDVQISPGGKYTCLAENEAGSQTSVGLIVKEPA 5668  
QY 144 ----- 143  
Db 5669 KIIRAEELIQVTAGDPATLEYTVAGTPELKPWKYKDRPLVASKKYRISFKNNVAQLKPY 5728  
QY 144 ----- 143  
Db 5729 SAELHDSQYTFEISNEVSGSSCETTFTVLDLDRDIAPFTKPLRNVDSVVNGTCRLDCKIA 5788  
QY 144 ----- 143  
Db 5789 GSLPMRVSWFKDQKEIAASDRYIAFVEGTASLEIIRVDMNDAGNFTCRATNSVSGKSS 5848  
QY 144 ----- 143  
Db 5849 GALLIVQEPATFVEKLEPSQLLKKGDATQLACKVTGTPPIKITWFANDREIKESSKHMSP 5908  
QY 144 ----- 143  
Db 5909 VESTAVLRLTDVGIEDSGBYMCEAQNEAGSDHCSIVIVKESPYTFKEKPIEVLKEYDV 5968  
QY 144 ----- 143  
Db 5969 MLLAEVAGTPPEITWFKONTILRSGRKYKFTIQDHLVSLQLKLVAAADAGYQCRVNE 6028  
QY 144 ----- 143  
Db 6029 VGSSICSAVTLREPATITEEAVSIDVTOGD PATLQVKFSGTKEITAKWFKDQBELTGS 6088  
QY 144 ----- 143

Db	6089	KYKISVTDVTSILKIISTEKDGEYTFEVQNDVGRSSCKARINVLDLII	PPSFTKKLKK	6148	Db	7169	PELSVEMYKDGKLLTSSQKHKFSFYNKISSILRILSVRODAGTYTFQVQNNVGKSSCTAV	7228
QY	144	-----	-----	143	QY	149	-----	148
Db	6149	MDSIKGSFIDLECIVAGSHPIQWPKDQDEISASEKYKFSFHDNTAFLEISQLEBPYPF	6208	Db	7229	VDVSDRAVPSPFRRLLKNTGGVLGASCILECKVAGSSPISVANWFHEKTKIVSGAKYQTTF	7288	
QY	144	-----	-----	143	QY	149	-----	148
Db	6209	VEKPOSQDVNPNTRVOLKALVGTA PMWTIKWFKDNKELHSGAARSVWKDDTSTLSLELFAA	6268	Db	7289	SDNVCTQLNSLSDSMGNYTCTVAANVAGSDECRAVLTVQEPSPSFVKPEPLEVLPKGNV	7348	
QY	144	-----	-----	143	QY	149	-----	148
Db	6269	KATDSGTIQLSNDVGTATSKATLFVKEPPQFIKKPSPVLVRNQSTTFEQINGTEPK	6328	Db	7349	TFTSVIRGTPPKVNWFRGARELVKGRCNIFYEDTVAELEFNIDI SOSGEYTCVVSNN	7408	
QY	144	-----	-----	143	QY	149	-----	148
Db	6329	IRVSWYLDGNEITAIQKHGIFSIDGLATFQISGARVENSGTYVCEARN DAGTASCIELK	6388	Db	7409	AGQASCITRLPVKGLLKWIHPFCFLYTDILVFFCLTAVLCHLLBPAAFLKELSDHSVEP	7468	
QY	144	-----	-----	143	QY	149	-----	160
Db	6389	VKEPPTFIRELKEVEVVKISDVELECEVTGTPPFVTLKNNREIRSSKKYTLTDRVSVF	6448	Db	7469	GKSIILESTYTGTLPISVTWKKDGFNITTEKCNIVTTEKTCILBILNSTKRDAGQYSCE	7528	
QY	144	-----	-----	143	QY	161	-----	160
Db	6449	NLHITKCDPSDTGEYQCI VSNEGGSCSTRVALKEPPSFIKKIENTTTLKSSATFQST	6508	Db	7529	IENEAGRDVCGALVSTLEPPYFVTEBLEPAAVGSVSLQCVAGTPEITVSWYKGDTKL	7588	
QY	144	-----	-----	143	QY	161	-----	168
Db	6509	VAGSPISITWLKDDQILDDEDDNVIYISFVDSVATLQIRSVDNHSGRYTCQAKNESGVFR	6568	Db	7589	RTPPYRYFTNNVATLVFNKVINDSGEYTCKAENSIGTASSKTIVFRIQERQLPPSPAR	7648	
QY	144	-----	-----	143	QY	169	-----	168
Db	6569	CYAFLLVQEPAPQIVKAKSVDVTEKDPMTLECVVAGTPELKVWLKDGKQIVPSRYFSMS	6628	Db	7649	QLKDIEQTVGLPVTLTCLRLNGSAPIQVCWYRQDGVLLRDDENLQTSFVDNVATLKILQTDL	7708	
QY	144	-----	-----	148	QY	169	-----	168
Db	6629	FENNVSFRIQSVWKQDSQGYTFKVENDFGSSCDAYLRVLDQNI	PPSFTKKLTMDKVL	6688	Db	7709	SHSGYSCASNPLGTASSARLTAREPKSPFFDIKPVSIDVIAGESADFECHVTGAOP	7768
QY	149	-----	-----	148	QY	169	-----	172
Db	6689	GSSIHECKVSGSLPISAQWFKDGKEISTSAKRYLVCHERSVSLEVNLEEDTANYTCK	6748	Db	7769	MRITWSDKNKEIRPGGNYITTCVGNTPHLRLILKVGKDSQGYTCQATNDVQKDMCSAQLS	7828	
QY	149	-----	-----	148	QY	173	-----	172
Db	6749	VSNVAGDDACSGILTWKEPPSFLVKPGRQQAIPDSTVEFKAILKGTPTPPKIKWFKDDVEL	6808	Db	7829	VKEPPKFKLEASKVAKQGESIQLECKISGPEIKVSWFRNDSSELHESWKYNMSPINSV	7888	
QY	149	-----	-----	148	QY	173	-----	172
Db	6809	VSGPKCFIGLGSTSFNLNLYSDASKTGQYTCHTVNDVGSDSCTTMLLVTEPPKFKVCKLE	6868	Db	7889	ALLTINEASAEDSGDYICEAHNGVGDASCSTALTVKAPPVFTQKPSVPGALKGSDVILOC	7948	
QY	149	-----	-----	148	QY	173	-----	172
Db	6869	ASKIVKAGDSSRLECKIAGSPEIRVWFRNEHELPA SDKYRMTFIDSVAVIQNNILSTED	6928	Db	7949	EISGTPPFEVWVKORKQVNSKKFKITSKHFDTSILHLNLEASDVGEYHCKATNEVGSD	8008	
QY	149	-----	-----	148	QY	173	-----	172
Db	6929	SGDFICEAQNPA GSTCSKVIIVKEPPVSPFPPIVETLKNAEVSLCELSGTPPFEVW	6988	Db	8009	TCSCSVKPEPPRFVKKLSDTSTLIGDAVELRAIVEGFQFISVWLKDRGEVIRESENTR	8068	
QY	149	-----	-----	148	QY	173	-----	172
Db	6989	YKDKRQLRSSKKYKIASKNPHTSIHILNVDTSIDIGEYHCKAQNEVGSDTCVCTVKLKEPP	7048	Db	8069	ISFIDNIATLQLGSPPEASNSGKYICOIKNDAGMRECSAVLTVLEPARIIEKPEPMTVTTG	8128	
QY	149	-----	-----	148	QY	173	-----	172
Db	7049	RFVSKLSLTVVAGEPAEQASIEGAQPIFVQWLKEKEEVIRESINIRITFVENVATLQF	7108	Db	8129	NPPALECVVTGTPELSAKWFKDRELSADSKHHITFINKVASLIKIPCAEMSDKGLYSFEV	8188	
QY	149	-----	-----	148	QY	173	-----	172
Db	7109	AKAEPANAGKYICOIKNDGGRENMATLMVLEPAVIVEKAGPMTVTVTGETCTLECKVAGT	7168	Db	8189	KNSVGKSNCTSVVHSDRIVPPFPFIRKLKDVNAILGASVLECRVSGSAPISVGWFQDGN	8248	
QY	149	-----	-----	148	QY	173	-----	172
Db	7169	EIVSGPKQSSFSSENVCTNLNLSLEPSDTGIYTCVAANVAGSDECSAVLTVQEPSPFEQT	8308	Db	8249	EIVSGPKQSSFSSENVCTNLNLSLEPSDTGIYTCVAANVAGSDECSAVLTVQEPSPFEQT	8308	



QY	173	-----	172
Db	8309	PDSVEVLPGMSLTFTTSVIRGTPPFVKWFKGRELVPGBSCNISLBDFFTELELFEVQPL	8368
QY	173	-----	172
Db	8369	ESGDYSCLVNDAGSASCTHLLFVKBPATFVKRLADFSVETGSPVILEATYTGTPPISVS	8428
QY	173	-----	172
Db	8429	WIKDEYLIQSSEKCSITMTEKSTILBILSTIEDYQAQYSLIENEAGQDICEALVSVLEP	8488
QY	173	-----	172
Db	8489	PYFIEPLEHVEAVIGBPATLQCKVDGTPBIRISWYKHTKLSAPAYKQKFNNAVSLVI	8548
QY	173	-----	172
Db	8549	NKVDHSDVGEYSCKADNSGVASSAVLVIKARKLPFFPARKKDVHETLGPVAFECRI	8608
QY	173	-----	172
Db	8609	NGSEPLQVSWYKDGVLKKDDANLQTSFVHVNATLQILQTDQSHIGQYNCASNPLGTASS	8668
QY	173	-----	172
Db	8669	SAKLILSEHEVPPFFDLKPVSVDLALGESGTFKCHVTGTAPIKITWAKDNREIRPGNYK	8728
QY	173	-----	172
Db	8729	MTLVENTATLTVLKVGKG DAGQVTCYASNIAGKSCSAQLGVQBPFRFIKKLEPSRIVKQ	8788
QY	173	-----	172
Db	8789	DEFTRYECKIGSGPEIKVLWYKDETELQESSKFRMSFVDSVAVLEMHNL SVEDSGDYTCE	8848
QY	173	-----	172
Db	8849	AHNAAGSASSTSLKVEPPIPRKKPHPIETLKGADVHLECELQGTTPFHVSVYKDKREL	8908
QY	173	-----	172
Db	8909	RSQKTKYIMSENFLTSIHILNDAADIGBYQCKATNDVGS DTCVGSIALKAPPRFVKCLS	8968
QY	173	-----	172
Db	8969	DISTVVGKEVQLQTTTIEGABPISVVMFKGGEIVRES DNIWISYENIATLQFSRVEPAS	9028
QY	173	-----	172
Db	9029	AGKYTCQIKNDAGMQECFATLSVLEPATIVEKPESIKVTTGDTCTLECTVAGTPPELSTKW	9088
QY	173	-----	172
Db	9089	FKDGKELTS DNKYISFFNKVSGLKIINVAPSDGVSYPEVQNPVGKDSCTASLQVSDRT	9148
QY	173	-----	172
Db	9149	VPPSFRKLKETNGLSGSSVVMCKVYGPPIPSVSWFHEGNEISSGRKYQTTLTDNTCAL	9208
QY	173	-----	172
Db	9209	TVMLESDSGDYTCIATNWAGSDECSAPLTVREPPSFVKQPDPMVDLGTNTVFTSIVK	9268
QY	173	-----	172
Db	9269	GTPPFSVWPKGSSSELVPGDRCNVSLSDSVAELELFDVDTQSGEYTCIVSNEAGKASCT	9328
QY	173	-----TF-----	174
Db	9329	THLYIKAPAKFVRLNDYSIEKGKPLILEGTFGTGTPPISVTVKKNGINVTPSQRCNITTT	9388

QY	175	-----	174
Db	9389	EKSAILEIPSTVEDAGQYNCYENASGKSDCSAQILILEQKLPPSFSQRLRDVQETVGL	9448
QY	175	-----	174
Db	9449	PVPFDCAISGSEPISVSWYKDGKPLKDSPNVQTSPLDNTATLNI FKTDRLAGQYSC TAT	9508
QY	175	-----NPMFFNV-----	181
Db	9509	NFIGSASSARLILTEGKNPPFFDIRLAPVDAVVGESADFECHVTGTGTQPIKVSMAKDSRE	9568
QY	182	-----	181
Db	9569	IRSGGKYQISYLENSAHLTVLKVDKDGSGQYTCYAVNEVGKDSCTAQLNIKERLIPPSFT	9628
QY	182	-----	181
Db	9629	KELSETVEBTEGNSFKLEGRVAGSQPI TVAWYKNNIEIQPTSNCEITFKNNTLVQLVRKA	9688
QY	182	-----	181
Db	9689	GMNDAGLYTCKVSDNAGSALCTSSIVIKEPKPPVDDQHLTPVTVSEGEYVQLSCHVOGS	9748
QY	182	-----	181
Db	9749	EPRIQWLKAGREIKPSDRCSFSPASGTAVLELRDVAKADSGDYVCKASNVAGSDTTKSK	9808
QY	182	-----	181
Db	9809	VTIKABPTQFTKRIQNI VVSEHQSATFECEVSFDDAIVTWYKGPTELTESQKNFRNDGR	9868
QY	182	-----	181
Db	9869	CHYMTIHNVTPDDEGVYSVIARLEPRGEARSTAEVLTTKETIKLELKPDDIPDSRVP IPT	9928
QY	182	-----	181
Db	9929	MPIRAVPPEEIPPVVAPPILLLPTPEKKPPPKRIEESHERKVPKAKVPEKKAPPPPKVI	9988
QY	182	-----	181
Db	9989	KKPVIEKIEKTSRRMEEEKVQTVKPEIKPAIPLPAPEPKPKEAGVPKKTPTSPIEAERR	10048
QY	182	-----	181
Db	10049	KLRPGSGKEPPDEAPFTYQLKAVPLKFVKEIKDILITSESEFVGSSAIFECVSPSTAIT	10108
QY	182	-----	181
Db	10109	TWMKDSNIRESPKHFADGDKRKLHIIDVQLSDAGEYTCVLRNLGNKEKTSTAKLVVEE	10168
QY	182	-----	181
Db	10169	LPVRFVKTLEEBVTVKGOPLYLSCELNKERDVWRKOGKI VVEKPGRIVPGVIGLMRAL	10228
QY	182	-----	181
Db	10229	TINDADDDTAGTYTTVTENANNLECSSCVKVVVEIRDWLKPIRDQHVKPKGTAFACDI	10288
QY	182	-----	181
Db	10289	AKDTPNIKWFGYDEIPABPNDKTEILRDGNHLYLKIKNAMPEDIAEYAVEIEGKRYPAK	10348
QY	182	-----	181
Db	10349	LTLGEREVELLKPIEDVTIYKESAFDAEISEADIPGQWKLGELLRSPSTCEIKAEGG	10408
QY	182	-----	181
Db	10409	KRFLTLHKVLDQAGEVLYQALNAITTAITVKEIELDFAVPLKDVTVPRROARFECVL	10468
QY	182	-----	181



QY	182	-----	181	QY	193	-----	198
Db	12689	ILKQDKRITINVPKSTVTIIVDSKRSDTGTIIIEAVNVCGRATAVVEVNVLDKGPAA	12748	Db	13769	IYEDVTSNMLYKWNPKDNGSPIILGYWLEKREVNSTHWSRVNKSLLNALKANVDGLLEG	13828
QY	182	-----	181	QY	199	-----	202
Db	12749	FDITDVNESCLLTWNPPRDDGSKITNVVVERRATDSEVWHKLSVTKDTNFKATKLIP	12808	Db	13829	LTYVFRVCAENAAFGPKSPSPDPKTAHDPIISPFGPPPIPRVTDTSSTTIELEWEPPAFNG	13888
QY	182	-----	181	QY	203	-----	205
Db	12809	NKEYIFRVAENMYGVEPVQASPIAKYQFDPGPPTRLEPSDITKDVAITLTWCEPDD	12868	Db	13889	GGEIVGYFVDKQLVGTNEWSRCTERKMIKVRQYTVKEIREGADYKLRVSAVNAAGEPPGE	13948
QY	182	-----	181	QY	206	-----	205
Db	12869	GGSPITGYWVERLDPTDKWVRCKMPVKDITTVRVKGLTNKKYRPRVLAENLAGPGKPS	12928	Db	13949	TQPTVAEPQEPAPVELDVSVKGGIQIMAGKTLRIPAVTVGRPVPTKWTKBEGELDKOR	14008
QY	182	-----	186	QY	206	-----	205
Db	12929	KSTEPILIKDPIDPPSPRRLWLEVINITKNTADLKWTVPBKDGSPITNVIYVEKRDVRRKG	12988	Db	14009	VVIDNVGTSKSELIIDALRKDHGRVYITATNSCGSKFAAARVEVPDVGPFVLDLKPVTN	14068
QY	187	-----	190	QY	206	-----	211
Db	12989	WQTVDTTVKDKCTVTPLTEGSLYFRVAENAIQSDYTEIEDSVLAKDTFTTPOPPYA	13048	Db	14069	RMWCLLNKSDPEDDGGSEITGFIIBERKDAKMTWRQPIETERSKCDITGLLEGQEKYKRV	14128
QY	191	-----	190	QY	212	-----	217
Db	13049	LAVVDVTKRVDLKWBPKNKDGGRPIQRYVIEKKERLGRWVKAGKTAGDCNFRVTDVI	13108	Db	14129	IAKNKFGCGPPVEIGFILAVIDPLGPPTSPLRTYTERTKSTITLDMKEPRNGSGSIOGY	14188
QY	191	-----	190	QY	218	-----	217
Db	13109	EGTEVOQVRAENAGVHPSEPTIELSDTPSPPLDLHVTDAGRKHAIAWKPE	13168	Db	14189	IIIEKRHDKDPFERNVNRKLCPTTSFLVENLDEHQMVEFRKAVNEIGSEPSLPLNVVIO	14248
QY	191	-----	190	QY	218	-----	217
Db	13169	KNGGPIIGHVEMCPGTEKWRVNSRPIKOLKPKVEGVPVDPKEYVLRVAVNAIGVS	13228	Db	14249	DDEVPTTIKLRLSVRGDTIKVKAGBPVHIPADVTGLPMPKIEWSKNETVIEKPTDALQIT	14308
QY	191	-----	190	QY	218	-----	217
Db	13229	EPSEISENVAKDPCKPTIDLEHDIIVIEGKLSIPVFRAPVPVTVSWHKDGKEVKA	13288	Db	14309	KEEVSREAKTELSPKAVREDKGTVTYTASNRLGVSFRNVHVYVDRPSPPRLAVTDI	14368
QY	191	-----	190	QY	218	-----	217
Db	13289	SORLTMKNDHI SAHLEVPKSVRADAGIYITITLENKLGSAFINVKVIGLPGCKDIKAS	13348	Db	14369	KAESCYLTDAPLDNGGSEITHYVIDKRDASRKKAEBEVTNTAVEKRYGIWKLIPNGQY	14428
QY	191	-----	190	QY	218	-----	217
Db	13349	DITKSSCKLTWBPPEFDGPTLHYVLRERAGRTYIIPVMSGENKLSWTVKDLIPNGEY	13408	Db	14429	EFRVAVNKYGISDECKSKVVIQDPYRLPGPPGPKVLARTKGSMLVSWTTPDLNNGGSP	14488
QY	191	-----	190	QY	218	-----	217
Db	13409	FFRVKAVNVGGGEYIELKNPVIAQDPKQPPDPVDVEHNPTAEMTITWKPLDYDGS	13468	Db	14489	ITGYWLEKREEGSPYWSRVSRAPITKVGLKGVEFNVPRLLLEGVKYQFRAMAINAAGIGPP	14548
QY	191	-----	190	QY	218	-----	217
Db	13469	KIMGYIIIEKIAKEERWKRNEHLVPLITYTAKLEBEGKEYQPRVRAENAAAGISESRAT	13528	Db	14549	SEPSDEVAGDPIFPFGPPSPCEVKDKTKSSISLWKPAPKDGGSPIKGIYIVEMQEGTT	14608
QY	191	-----	190	QY	218	-----	217
Db	13529	PPTKAVDPIDAPKVIILRTSLEVKRGDEIALDASISGSPYTTIWKDENVIVPEEIKKRA	13588	Db	14609	DMKRVNEPDKLITTCCECVVPNLKLRYRFRVKAVNEAGESEPSDTTGEIPATDIOEPE	14668
QY	191	-----	190	QY	218	-----	217
Db	13589	APLVRERKEVQEEFPVLPLOTQLSDNSKKGESQLRVDRDSLDPDHGLYMIKVENDHGI	13648	Db	14669	VFIDIGADCLVKAGSQIRIPAVIKGRPTPKSSWEFDGKAKMKLETAENSVIIIBE	14728
QY	191	-----	190	QY	218	-----	219
Db	13649	AKAPCTVSLDTPGPPINFVEDIRKTSVLCKWEPLDDGGSEIINYLEKKDKTKPDSE	13708	Db	14729	CKRSHTGKYSITAKNAGOKTANCRVKVMDVPFPKDLKVSDITRGSCHLSWMPDDGG	14788
QY	191	-----	192	QY	220	-----	219
Db	13709	WIVVTSLRHCKYSVTKLIEGKEYLFRVRAENRFGPGPCVKPLVAKOPFGPPADAPKP	13768	Db	14789	DRIKGVIEKRTIDGKAMTKVNPDCGSTTFVVPDLLSEQQYFFRVRAENRFGIGPPVETI	14848
				QY	220	-----	229

Db	14849	QRTTARDPIYPPDPPIKIKIGLITKNTVHLSWPKPKNDGSPVTH--YIVECLAWDPTGT	14906
Qy	230	-----	229
Db	14907	KKEAWROCNKRDVBELOFTVEDLVEGGEYFRVKAUNAAGVSKPSATVGPVTVKDQCPP	14966
Qy	230	-----	229
Db	14967	SIDLKEFMEVEEGTDVNI VAKI KGVPPPTLTWFKAPPKPKDNKEPVLVDTHVKNLVDDT	15026
Qy	230	-----	229
Db	15027	CTLVIQSRRSRDTGLYITAVNNLGTASKEMRLNLGRPPGPVPIKFESVSADOMTSLW	15086
Qy	230	-----	229
Db	15087	FPPKDDGSKITNVIVEKREANRKTWHVSSPEKECTYTTPKLLGHEHYFRIMAQNKYG	15146
Qy	230	-----SRNLFRVP	237
Db	15147	IGEPDSEPETARNLFSVPGAPDKPTVSSVTRNSMTVNWEEPEYDGGSPVTGYWLEKDT	15206
Qy	238	-----	237
Db	15207	TSKRWRKVRNRPDKAMTLGVSXKVTGLIEGSDYQFRVYAINAAGVGPASLPSPATARDP	15266
Qy	238	-----KYINGTKL-	245
Db	15267	IVTDWTKSSADLEWSPLPKDGGSKVTGYIIVEYKEEGKEWEKDKVEVRGTKLVVTGLKE	15326
Qy	246	-----	245
Db	15327	GAFYKFRVRAVNIAGIGEPGEVTDVIEMKDRLVSPDLQLDASVRDRIVWHAGGVIRIAY	15386
Qy	246	-----	245
Db	15387	VSGKPPPTVTWNNERTLPOEATIEETAISSSMVKNQCORSHQGVYSLAKNEAGERKKT	15446
Qy	246	-----KNT-	248
Db	15447	IIVDVLAVPERPDELVKEVTKNTVLTWNPYPKYDGGSEIINVLBSRLIGTEFKHKVTN	15506
Qy	249	-----	248
Db	15507	DNLLSRKYTVKGLKEGDTYEVRSVAVNI VQGKPSFCTKPIITCKDBLKLITIRVGQAFAL	15566
Qy	249	-----MRKJKEKQ-	256
Db	15567	TGRYSGPKPKVSWFKDEADVLDDRTHIKTTPATLALKEKIKAKRSDSGKYCVVVENSTG	15626
Qy	257	-----	256
Db	15627	SRKGFQVNVVDHPGPPGVSPFDEVTKDVMVLSWKPPPLDDGSKITNYIIIEKKEVGKV	15686
Qy	257	-----	256
Db	15687	WMPVTSASAKTCKVSKLLEGKDYI FRIHAENLYGISDPLVSDSMKAKDRFRVPADQP	15746
Qy	257	-----	256
Db	15747	IVTEVTKDSALVTWPKHDGKPIITNYIIIEKRETMKRWARTKDPHPTKFRVPDDLLE	15806
Qy	257	-----	256
Db	15807	GCQYEFVRAENEIGIGDPPSPKPVFAKDPFAKPSPPVNPPEAIDTTCNSVDLTWQPPRH	15866
Qy	257	-----	256
Db	15867	DGGSKILGYIIVEYQKVGDSEWRANHTPSCBETKYKVTGLRDGQTYKFRVLAVNAAGES	15926
Qy	257	-----	256

---

Db	15927	DPAHVPEPVLVKDRLEBPPELLIDANNAREQHIKVGDTRLRLSALIKGVPPFKVTWKEDRD	15986
Qy	257	AP-----	258
Db	15987	APTAKRIDVTPVGSKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLVDKPGPPRDLEV	16046
Qy	259	-----	258
Db	16047	SEIRKDSCYLTWKEPLDDGGSVITNVVERRDVASAQNWSPLSATSKKSHFPAKHLNEGNO	16106
Qy	259	-----	258
Db	16107	YLFRAAENQYGRGPPFVETPKPIKALDPLHPGPKDLHHVDVDKTEVSLVWNKPRDGG	16166
Qy	259	-----	258
Db	16167	SPITGYLVBEGTODWIKFTVTNLCVVTGLQQGKTYRFRVKAENIVLGLSPDTTIP	16226
Qy	259	-----	258
Db	16227	IECQEKLVPPSVELDVKLI EGLVVRKAGTTRPAAIIRGVVPVPTAKWTTDGSSEIKTDEHYT	16286
Qy	259	-----	258
Db	16287	VETDNFSSVLTIKNCLRRDTGEYQITVSNAGSKTVAVHLTVLDPGPTGPINILDVTP	16346
Qy	259	-----	258
Db	16347	EHWTSWQPPKDDGGSVINYI VEQDTRKOTGWVSVSSGSKTKLKI PHLQKGEYFVR	16406
Qy	259	-----	258
Db	16407	RAENKIGVPLDSTPTVAKHKFSPSPPGPKPVVTDITENAATVSWTLPKSDGSGSPITY	16466
Qy	259	-----	258
Db	16467	YMERREVTGKVRVVKNTPIADILKFRVTGLYEGNTYEFYRFAENLAGLSKSPSSDPKAK	16526
Qy	259	-----	258
Db	16527	RPIKPPGPINPKLKDKSRETADLVWTKPLSDGSGPILGYVVECKPGTAQWRINKDEL	16586
Qy	259	-----	258
Db	16587	IRQCAFVPLTEGNEYRFRKAAINIVEGEPELAESVIAKDILHHPPEVELDVTCDVI	16646
Qy	259	-----	258
Db	16647	TVRVQOTIRILARVKRPEPDI TWTKGKVLVREKRVDLI ODLPRVELQI KEAVRADHGK	16706
Qy	259	-----	258
Db	16707	YIISAKNSSGHAQSAI VNVLD RPPCQNLKVTVNTKENTISWENPLDNGSGSEITNFIV	16766
Qy	259	-----	258
Db	16767	EYRKPNQGSIVASDVTKRLIKANLLANNETFRVCAENKVGVGPTIETKTPILAINPI	16826
Qy	259	-----	258
Db	16827	DRPGEPENLHIADKGTFFYLVKWRPPDYDGGSPNLSYHVERRLKGSDDWVERVHKSGIKET	16886
Qy	259	-----VKEQFEK-	265
Db	16887	HMVDRCVENQIYEFVQTKNEGGSWMVKTEEVVVVEDLQKPVLDLKL SGLVTVKAGDT	16946
Qy	266	-----KAK-	268
Db	16947	IRLEAGVRGKPPFPEVAVTKDADTLTRS PRVKIDTRADSSKSLTKAKRSGGKVVTA	17006
Qy	269	-----	268
Db	17007	TNTAGSFVAYATVNVLDKPGPVRNLKIVDVSSDRCTVCWDPDDGGCEIQNYILEKCT	17066

QY	269	-----	268	QY	290	-----	289
Db	17067	KRMVSTYSATVLTGTTVRLIEGNEYIFRVAENKIGTGPPTESKPVIAKTKYDKPGR	17126	Db	18147	GLQEGTEYFRVTAINKAGPKSPDASKAAYARDPQYPPGPAPPKYVDTRSSVLSWG	18206
QY	269	-----	268	QY	290	-----	289
Db	17127	PDPEVTKVSKEEMTVVWNPPEYDGGKSTGYFLEKKEKSHTRWVPVNSAIPERRMKVQ	17186	Db	18207	KPAYDGGSPIIIGYLVEVKADSDNWVRCNLQNLQKTRFEVTLGMDTOYQPRVAVANKI	18266
QY	269	-----	268	QY	290	-----	289
Db	17187	NLLPDHEYQPRVKAENEIGIGBPSLPSRPVAKDPIEPGPPPTNFRVVDTTKHSITLGMG	17246	Db	18267	GYSDPSPDVKHYPKDIILIPPEGELDADLRKTLILRAGVTMRLYVPVKGPRPPKITWSKP	18326
QY	269	-----	268	QY	290	-----	289
Db	17247	KPVYDGGAPIIGYVVENRPKIADASPDEGMRKRCNAAQLVRKEFTVTSLDENOEYFRMS	17306	Db	18327	NVNLDRIGLDIKSTDFTLRCENVNKYDAGKYLITLNSCGKKEYTIVVKVLDTPGPP	18386
QY	269	-----	276	QY	290	TNVT-----	293
Db	17307	LKHGHVSNAPPENDGSGQVTHYVEKREADRTWSTVTPDEVKTSFHVTNLVPGNYYF	17366	Db	18387	VNVTVKEISKDSAYVTWEPPIIDGSGPIINYVVKQDAERKSWSTVTECSKTSFRVPLN	18446
QY	277	-----	276	QY	294	-----	293
Db	17367	RVTAVNEYGVPTDVPKPLASDPLSEPPPRKLEVTMTKNSATLWMLPPLRDCGAKI	17426	Db	18447	EEGKSYFRVFAENEYIGDPCGETRDAVKASOTPGPVVDLKVRSVSKSCSIGWKPHSD	18506
QY	277	-----	276	QY	294	-----	297
Db	17427	DGYITSYREEQPADRWTEYSVVVKDLSLVVTGLKEGKYKFRVAARNAVGVSIPREAGV	17486	Db	18507	GGSRIGYVVDVFLTEENKQVRVMSLSQYSAKDLTEGKEYTFRVSAENENGEGTPTSEIT	18566
QY	277	-----	276	QY	298	TAAR-----	301
Db	17487	YEAKEQLPPKILMPQIITIKAGKURIEAHVYGKPHPTCKWKKGDEVVVTSHLAVHKA	17546	Db	18567	VVARDDVAPDLDLKGLPDLCYLAKENSFRLKPIKIGKPAVSVMKKGEDPLATDTRVS	18626
QY	277	-----	281	QY	302	-----	301
Db	17547	DSSSILIIKDVRKDSGYLSLTAESENSGTDQKIKVVMDAPGPPQPPDISIDADACS	17606	Db	18627	VESSAVNTLIVVDCQKSDAGKYITILKNVAGTKEGTISIKVVKGPIPTGPIKFDVTA	18686
QY	282	-----	281	QY	302	-----	301
Db	17607	LSWHIPLEGGSNITNYIVEKCDVSRGDWVTALASVTKTSRVGKLIPGOEYIFRVAEN	17666	Db	18687	EAMTLKMAPKDDGGSEITNYILEKRDVNNKVVTCASAVQKTTFRVTRLHEGMEYTFRV	18746
QY	282	-----	281	QY	302	-----	301
Db	17667	RFGISEPLTSPKMAVQPPFGVPSEPKNARVTKNKDCIFVWDRPDSGGSPIIIGYLIER	17726	Db	18747	SAENKYGVEGLKSEPIVARHFDVDPDAPPNNIVDVRHDSVSLTWDPKTKGGSITGY	18806
QY	282	-----	281	QY	302	-----	301
Db	17727	KERNLLVWKANDTLVRLTEYPCAGLVEGLEYSFRIYALNKAGSSPPSKPTEYVTARMPV	17786	Db	18807	HLEFKERNLLWKRAKTFPIRMRDPKVTGLTEGLEYEFRVMAINLAGVKGKPSLSPBVVA	18866
QY	282	-----	281	QY	302	-----	301
Db	17787	DPGKPEVIDVTKSTVLIWARKPHDGGSKIIIGYFVEACKLPGDKWRCNTAPHQIPQEE	17846	Db	18867	LDPTEAPTIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGKDIRPSDITQIT	18926
QY	282	-----	281	QY	302	-----	301
Db	17847	YTATGLEEKAQYQFRAIARTAVNISPSPSDPVTILAENVPPRIDLSVAMKSLLTVKAG	17906	Db	18927	STPTSSMLTIKYATRKADAGEYITATNPFGTKVEHVKVTVLDPVPGPPGVEISNVSAEKA	18986
QY	282	-----	289	QY	302	-----	304
Db	17907	TNVCIDATVPGKPMPTVSWKKDGTLLKPAEGIKMAQORNICTLELFSVNRKDSGDYITA	17966	Db	18987	TLTWTTPLEDGGSPKSYILEKRETSRLWTVVSEDIQSCRHVATKLIQGNEYIFRVSAV	19046
QY	290	-----	289	QY	305	-----	312
Db	17967	ENSSGKSATIKLVLDKPGPPASVKINKMYSDRAMLSWEPPLEDGGSEITNYIVDKRET	18026	Db	19047	NHYGKEPVQSEPVKMWDRFGPPGPEKPEVSNVTNTATVSMKRFVDDGGSEITGYHVE	19106
QY	290	-----	289	QY	313	-----	312
Db	18027	SRENWAQVSATVITSCSVEKLIEGHEYQFICAENKYGVDVFTPEAIKPNYPDPGR	18086	Db	19107	RREKSLRWRAIKTPVSLRCKVTGLQEGSYEFVRVSAENRAGIOPPNPPGPPSPHVT	19166
QY	290	-----	289	QY	313	-----	326
Db	18087	CDPPVISNITKDHMTVSWKPPADDGSGPITGYLLEKRETOQVNWTKVNRKPIERTLKAT	18146	Db	19167	DTTKKSASLAWGKPHYDGGLEITGYVVEHQVKGDEAWIKDITCTALRITQFVVVPDLQTK	19226
				QY	327	-----	326

Db	19227	KYNFRISAINDAGVGEPAV	102307	CCTURLFPVIKGRPAPEVK	20366
Qy	327	-----	340	-----	339
Db	19287	RPAPEVTWTKDNINLKRAN	20367	NSSGSKSAFVNVRLDTPGP	20426
Qy	327	-----	340	-----	339
Db	19347	RVLDTGCVLNLRPTDITKDS	20427	RKAYSTVATNCHTKSWKVD	20486
Qy	327	-----	340	-----	339
Db	19407	KCTYKVTGLSEGCEYFRVMA	20487	ITLMDVTRNSVLSWEKPEH	20546
Qy	327	-----	340	-----	339
Db	19467	LAWPKPHDGGSKITGYVIEA	20547	BEYSFRVSAQNEKGISDPR	20606
Qy	327	-----	340	-----	339
Db	19527	AGRSAPRESRPVIVKESQTM	20607	PTPAVTWHKDNVPLKQTTR	20666
Qy	327	-----	340	-----	339
Db	19587	DQILKQTVRNPETTATSTIL	20667	VIVLDKPGPTGPVKMDEV	20726
Qy	327	-----	340	-----	339
Db	19647	TGPIKFDEVSSDFVTFSWDP	20727	VARTTIKACRLKTGCEYQF	20786
Qy	333	-----	340	-----	339
Db	19707	LTTGLEVOFRVKAQNRVGVG	20787	DSMEVQWNEPISDGGSRV	20846
Qy	333	-----	340	-----	339
Db	19767	AENMAGSKSPSEPMALDPI	20847	VSAENIVGIGKPSVSECVA	20906
Qy	333	-----	340	-----	339
Db	19827	KRDLPNGRWLKANFSNILEN	20907	VEKKELPEGRWMAKSFNII	20966
Qy	333	-----	340	-----	339
Db	19887	DVEAPKIKVDVKFDTVILKA	20967	RDEVDPRIKMDPKYKTIIV	21026
Qy	333	-----	340	-----	339
Db	19947	FSTNLVNKDSRTRDSGAYTL	21027	TDFATSLSVKDAVRVDSGN	21086
Qy	333	-----	340	-----	339
Db	20007	SWFPPLDDGGAKIDHYIVQ	21087	TLAWKPLODGGSDIINVI	21146
Qy	333	-----	340	-----	343
Db	20067	YGVCPELESEPLAVNPYGP	21147	NKYGCEPLESEPVAKNPF	21206
Qy	333	-----	344	-----	346
Db	20127	DKAQRMWIKCNKKTLDLRY	21207	KRDKEGIRWTRCHKRLIG	21266
Qy	340	-----	347	-----	346
Db	20187	KPGPPGNPRVLDTSRSSIS	21267	IYKPGPPNNPKVIDITRSS	21326
Qy	340	-----	347	-----	346
Db	20247	SYTTITLTENQYKIRIYAMN	21327	KTNIEVKLEKHEYNFRICA	21386
Qy	340	-----	347	-----	346
			21387	RAGGSURLFVPIKGRPTPE	21446

QY	347	----	LYY-	381
Db	21447	LENSGTSAPVTVRLDTPSPVNLKVTEITKDSVITWEPPLDGGSKIKNIYVEKRE		22527
QY	347	----	----	384
Db	21507	ATRKSAVVVNTCHNSWKIDQLQEGCSYFRVTAENYIGIGLPAQTADPIKVAEVPQPP	----	22587
QY	347	-----SRNRTAVS-	----	384
Db	21567	GKTVDDVTNSVLSWTKEPHDGGSKIIQYIVEMQAKHSEKSECARVKSLOAVITNLT	----	22647
QY	355	-----EFM-	----	384
Db	21627	QGBEYLFRRVAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAPSSYSVQVQDLKIEVPIS	----	22707
QY	358	-----KNTH-	----	384
Db	21687	GRPKPTITWKGPLKQTTTRINVTDSLDTLISIKETHKDDGGQYIIVANVVGOKTAS	----	22767
QY	362	-----	----	384
Db	21747	IEIVTLKDDPPKGPVKFDVSAESITLSWNPLYTGGCQITNYIVQKRDTTTTVWDVVS	----	22827
QY	362	-----	----	384
Db	21807	ATVARTTLKVKLTGTEYQFRIFAENRYGQSFALSDPIVAQYPYKEPGPGKASKNSE	----	22887
QY	362	-----	----	387
Db	21867	CYVARDPCDPGTPPEIMVKRNEITLQWTKPVYDGGSMITGYIVEKRDLPDGRWMAKSF	----	22947
QY	362	-----	----	387
Db	21927	NVETQFTVSGLTQDORYEYFRVIKNAAGAIKSPSDSTGPITAKDEVELPRISMDPKFRD	----	23007
QY	362	-----	----	387
Db	21987	TIVNAGETFRLEADVHGKPLPTIEWLRGDKEIESARCEIKNTDPKALLIVKDAIRIDG	----	23067
QY	362	-----	----	387
Db	22047	GQYILRASVAGSKPFPVNVKLDLDPGPPGVPQVTGVTSEKSLTWSPLQDGGSDISH	----	23127
QY	362	-----VLJR 365	----	387
Db	22107	YVVEKRETSRLAWTVVASEVVTNSLKVTKLLEGNEYVFRIMAVNKYGVGEPLESAPVLMK	----	23187
QY	366	N-----	----	387
Db	22167	NPPVLPGPSLEVTNIAKDSMTVCWNRPDSDGSEIIGYIVEKDRSGIRWIKCNKRRI	----	23247
QY	367	-----ETPYTY-	----	387
Db	22227	TDLRLVTLGTEDHEHYEYFRVAENAGVGEPSATVYKACDPVKPGPPTNAHIVDTTK	----	23307
QY	374	-----	----	387
Db	22287	NSITLAWKPIYDGGSEILGYVVEICKADEEEMQIVTPQGLRVTRFEISKLEHQEYKI	----	23367
QY	374	-----GT-----LDMSS-	----	387
Db	22347	RVCALNKVGLGEATSPGTVKPBEDKLEAPELDDSELRGIVVRAGGSARIHIPKGRPT	----	23427
QY	381	-----	----	387
Db	22407	PEITWSREGEFTDKVQIEKGVNVTQLSIDNCDRNDAGKYILKLENSGSKSAFVTVKVL	----	23487
QY	381	-----	----	387
Db	22467	DTPGPPONLAVKEVRKDSAFVWEPPSLMEQORSSNYVIDKPKAYANVSSKCSKTSFKVE	----	23547
QY	387	-----	----	387

Db	23607	RIENLOGCSYYFVRVLASNEYIGLPAETTEPVKVSEPLPPGRGRTLVDVTRTATIKWE	23666	Db	24687	WOQPAFDGSGKITGYIVERDLPDGRWTKASTNTVTETQFIISGLTQNSQYERFVAFARNA	24746
QY	387	-----	386	QY	414	-----	413
Db	23667	KPESDGGSKI TGYVEMQTKGSEKSTCTQVKTLEATISGLTAGEBYVFRVAANEKGRS	23726	Db	24747	VCSI NSPSEVVGPI TCIDSYGGPVIDLPLEYTEVVKYRAGTSVKLRAGISGKPAPTI EWY	24806
QY	387	-----	386	QY	414	-----	413
Db	23727	DPQLGVPVIARDIEIKPSVELPFHTFNVKAREQLXIDVPFKGRPOATVNRKDGOTLKE	23786	Db	24807	KODKELQTNALVCVENTTDLASILIKDADRLNSGCVELKLRNAMGSASATIRVQLDKPG	24866
QY	387	-----	386	QY	414	-----	413
Db	23787	TTRVNVSSKTVTSLSIKEASKEDVGTIELCVSNSAGSITVPITIIILDRPGPGPIRID	23846	Db	24867	PPGGPIEFKVTVAEKITLLWRPPADGGAKITHYIVREKRETSRVVMSVSEHLEECIIT	24926
QY	387	-----	386	QY	414	-----	413
Db	23847	EVSCDSITISWNPPEYDGGCQISNYIIVEKKETTSTTWHIVSQAVARTSIKIVRLTTGSEY	23906	Db	24927	TKLI IKNEVI FRVRAVNKYGICEPLESDSVAKNAFVTPGPGPIPEVTKITKNSMTVWS	24986
QY	387	-----	386	QY	414	-----	413
Db	23907	QFRVCAENRYKSSYSESSAVVAEYFPSPGPGTKVHVHATKSTMLVTWQVPVNDGGR	23966	Db	24987	RPIADGGS DISGYFLEKRDKKSGLGWFKVLKETIRDTROKVTGLTENS DYQYRVCAVNAAG	25046
QY	387	-----	386	QY	414	-----	416
Db	23967	VIGYHLEYKERSILWSKANKILIA DTOMKVSGLDEGLMVEYRYVAENIAGICKSKCE	24026	Db	25047	QGFSEPSSEFYKAADPIDPDPGPPAKIRIADSTKSSITLWNSKPVVDGGSVAVTGYVVEIRQ	25106
QY	387	-----	386	QY	417	-----	416
Db	24027	PVPARDPCDPGQPEVTNITRKS VSLKWSKPHYDGGAKITGYIIVERREL PDGRWLKCNVT	24086	Db	25107	GEEB EWTTVSTKGEVRTTEYVVVSNL KPGVNY YFRVS AVNACAGQGEPIENNEPVQAKDILE	25166
QY	387	-----	386	QY	417	-----	416
Db	24087	NIQETFEVTELTEQRYEFRVAFARNAADSVSEPSHSTGPIIVKDDVEPRVMMDVKFRD	24146	Db	25167	AP EIDL DVALRTSVIAKAGEDVQLIPFKGRPPPTVTRKDEKNLGS DARYSIENTDSSS	25226
QY	387	-----	386	QY	417	-----	419
Db	24147	VI VVKAGEVLKINADIAGRPLPVISWAKDGI EIERARTEIISTDNHLLITVKDCIRRD	24206	Db	25227	LLTIPQVTRNDTGKXILTIENGVBPKSSTVSVKVLDTPAACQKQVHKVSRGTVTLWD	25286
QY	387	-----	386	QY	420	-----	419
Db	24207	GQVYLT LKNVAGTRSAVNCKVLDKPGPPAGPLEINGLTA EKCSLSWGRPDGGADIDY	24266	Db	25287	PPLIDGSGPI INYVIEKRDATKRTWSVVSHKCSSTSFKLIDLSEKTPFFRVLAE NEIGI	25346
QY	387	-----	386	QY	420	-----	419
Db	24267	YI VEQKPSPEKLGVTISISKDSVSLTLWKPEHDGGSRI VHYVVALEKQKNWKCAVA	24326	Db	25347	GBPCETTEPVKAAVPAPIRDL SMDSTKTSVILSWTKDPDGGSVITEYVVERKGGKRG	25406
QY	387	-----	392	QY	420	-----	419
Db	24327	KSTHHVVSGLRENSEYFFRVFAENQAGLSDPRELLLPVLIKEQLEPPEIDMKNFP SHTVY	24386	Db	25407	TWSHAGISKTCEIEVSQ LKQSVLEFRVPAKNEKGLSDPVTIGPITVKELIITPEVDLSD	25466
QY	393	-----	392	QY	420	-----	419
Db	24387	VRAGSNLKV DIPISGKPLPKVLTLSRGGVPLKATMRFNTEITAE NLTINLKS VTA DAGRY	24446	Db	25467	IPGAQVTVRIGHNVHLELPYKGP KPSISWLX DGLPK ESEFVRFSKTENKITLSIKNAK	25526
QY	393	-----	405	QY	420	-----	419
Db	24447	EITAANS GGTAKAFINIVLDRPGPTGPVWISDITEESVTLKWP KPYDGGQVNYIL	24506	Db	25527	KEHGKGYT VILDNAVCR IAVPITVITLGPSPKGP IREFDEIKADS VILSWDVPEDNGG	25586
QY	406	-----	405	QY	420	-----	419
Db	24507	LKRETSTAVWTEVSATVARTMMKVMKLT TGGEYQFRIKAENRFGISDHIDSACVTVKLPY	24566	Db	25587	EITCYSIEKRETSQTNWKMVCSSVARTTPKPNLVKDAEYQFRVRAENRYGVSQPLVSSI	25646
QY	406	-----	405	QY	420	-----	419
Db	24567	TTGPPSPTPWVNTVRESITVGHHEPVSGGS AVGVHLEMKDRNSILWOKANKL VIRT	24626	Db	25647	IVAKHQFRIPGPGKPEVIY NVTSDGMSLTWDAPVYDGGSEVTFGHVEKKERN SILWQKN	25706
QY	406	-----	SMG 408	QY	420	-----	419
Db	24627	HFKVTTISAGLIYEF RYVAENAGVKPSHPSEPVLAIDACEPPRNVRTIDISKNSVLS	24686	Db	25707	TSPISGREYRATCLVEGLDYQFRVYAENAGLS SPSDPKFTLAVSPVDPGTPDYIDVT	25766
QY	409	FO-----	RTF-----	QY	420	-----	419
				Db	25767	RETITLKNPPLRDGGSKIVGY SIEKROGNERWRCNFTD VSECQYTVTGLSPGDRYEPR	25826



QY	420	-----YDSLL-----	425
Db	25827	IIARNAVGTISPPSQSGIIIMTRDENVPPIVEFGPEYFDGLIITKSGBSLRIKALVOGRPV	25886
QY	426	-----	425
Db	25887	PRVTWPKDGVIEIKRMMEITDVLGSTSLFVRDATRDHRGVTYVEAKNAGSGAKAEIKVK	25946
QY	426	-----	425
Db	25947	VQDTPGVVGPPIRFTNITGKMTLWDAPLNDGCAPIITHYIEKRETSRLAWALIEDKCE	26006
QY	426	-----	425
Db	26007	AQSYTAIKLINGNEYQFRVAVNKGFGVGRPLDSPVVAQIQYTPDAPGIPESNITGNS	26066
QY	426	-----	425
Db	26067	ITLTWARPESDGGSEIQYILERREKSTRWVKVISKRPISETRPFKVTGLTEGNEYEFHV	26126
QY	426	-----	425
Db	26127	MAENAGVGSPASGISRLIKCEPVPNPPTTVVKVTDTSKTTVSLSKSPVFDGMEIIG	26186
QY	426	-----	425
Db	26187	YIEMCKADIADWHKVAEACVKTRYTVTDLOAGEYKFRVSAINGAGKGDSCVETGTIK	26246
QY	426	-----	425
Db	26247	AVDRLTAPELDIDANFKQTHVVRAGASIRLFIAYQGRPTPTAVWSKPDNSLSURADIHT	26306
QY	426	-----	425
Db	26307	DSFSTLTVCNCRNDAGKTYLTVENNSGSKIITFTVKVLDTPGPGPITFKDVTGSAITL	26366
QY	426	-----	425
Db	26367	MWDAPLLDGGARIHHYVVKREASRSWQVISEKTRQIPKVNDLAEGVPPYFRVSAVNE	26426
QY	426	-----	425
Db	26427	YGVGEPEMPEPIVATEQAPPRRLDVDTSKSSAVLWLPDHDGGSRTGYLLEMRQK	26486
QY	426	-----	425
Db	26487	GSDFWVEAGHTKQLTFTVERLVEKTEYEFVAKANDAGYSEPREAFSSVIIKEPQIEPTA	26546
QY	426	-----	425
Db	26547	DLTGITNQLITCKAGSPFTIDVPISGRPAPKVTWLEEMRLKETDRVSITTTKDRTLTV	26606
QY	426	-----	425
Db	26607	KOSMRDSCRYFLTLENTAGVKTFSVTVVVIGRPGPVTGPBIEVSSVSAESCVLSWGEPKD	26666
QY	426	-----	425
Db	26667	GGGTEITNYIVEKRESGTTAWQLVNSVVKRTQIKVTHLTKYMEYSPRVSSNRFVGSKPL	26726
QY	426	-----	425
Db	26727	ESAPIIAEHPFPVPPAPTREVYVHSANAMSIRWEEPYHDGSKIIGYWEKKERTILW	26786
QY	426	-----	425
Db	26787	VKENVPCLCNCYKVTGLVEGLEYPQRTYALNAGVSKASEASRPIMAQNVPDAPGRPEV	26846
QY	426	-----	425
Db	26847	TDVTRSTVSLNSAPAYDGGSKVVGVIIBERKPVSEVGDGRWLKCNNTIIVSDNFFTVALS	26906
QY	426	-----	425
QY	426	-----	425
Db	26907	EGDTYFRVLAKNAAGVISKSESTGVTCDREYAPPAKELDARLHGLDVTIRAGSDLVL	26966
QY	426	-----	425
Db	26967	DAAVGKPEPKIIWTGDKELDCEKVSLOYTKRATAVIKPCDRSDSGKYTLTVKNASG	27026
QY	426	-----	425
Db	27027	TKAVSMVKVLDSPGPGKLTVSRTQEKCTLAWSLPQEDGGAEITHYIVERRETSRLNW	27086
QY	426	-----	425
Db	27087	VIVEGECPTLSYVVTRLIKNNEYIFRVAVNKYGPGVVESEPIVARNSTFIPSGPIPE	27146
QY	426	-----	425
Db	27147	EVGTGKEHIIIOWTKPESDGGNEISNYLVDKREKSLRWTRVNKDYYVVDYTRUKVTSLME	27206
QY	426	-----	425
Db	27207	GCDYQFRVAVNAAGNSEPSEASNPFISCREPSYTPGPPSAPRVVDTTKHSISLAWTKPMY	27266
QY	426	-----	425
Db	27267	DGCTDIVGYVLEMOEKDQDQWYRVHTNATIRNTEFTVPDLKMGQKYSFRVAAVNVKGMSE	27326
QY	426	-----	425
Db	27327	YSESIAEIPVERIEIPDLELADDLKKTVTIRAGASLRMLVMSVSGRPPPVITWSKQIDL	27386
QY	426	-----	425
Db	27387	ASRAIIDTTESYSLLIIVDKVNYDAGKYTIEAENQSGKKSATVLVKVYDTPGPCPSVKVK	27446
QY	426	-----	425
Db	27447	EVSRSVTTITBIPTIDGAPVNNYIVEKREAAWRAFKTVTTKCKSKTYRISGLVBGTMY	27506
QY	426	-----	425
Db	27507	YFRVLPENIYGIBPCETSDAVLSEVPLVPAKLEVVDVTSTVLAWSKPLYDGGSRILT	27566
QY	426	-----	425
Db	27567	GYVLEACKAGTERWKKVTLKPTVLEHTVTSLSNEGEQYLFRIQAQNEKGVSEPREVTAV	27626
QY	426	-----	425
Db	27627	TVQDLRESERVTVETHTKVAKLTIRETTIRDTGEYTLKQNTGTSTETIKVILDKPGP	27686
QY	426	-----	425
Db	27687	PTGPIKIDEIDATISISWEPPELDGAPLSGYVVEQDRAHPGWLVPVSESSTRFKFT	27746
QY	426	-----	425
Db	27747	RLTEGNEYFRVAATNRFIGISYLOSEVIECRSSIRIPGPPETLQIFDVSRDGMTLTWYP	27806
QY	426	-----	425
Db	27807	PEDDGSQVQYIIVERKEVRADRWVRVNVKVPVTMTYRSTRGLTEGLEYEHRVTAINARGS	27866
QY	426	-----	425
Db	27867	GKPSRPSKPIVAMDPIAPPKQNPONPRVTDTRTSVLSAWSVPEDEGSKVTGYLIBMQKV	27926
QY	426	-----	425
Db	27927	DOHEWTKCNTTPTKIREYTLTHLPOGAEYFRVLACNAGGPGEPABVPQTVKTEMLEYP	27986
QY	426	-----	425

Db	27987	DYELDERYQEGIFVRQGGVIRLTPIPIKGRFPPICKWTKEGQDISKRAMIATSEHTELVI	28046	Db	29067	KEDKTRAMNYDEVEDTREVSMTKASHSTKELYKYMIAEDLGRGEFGIVHRCVETSSK	29126
QY	426	-----	425	QY	437	-----	436
Db	28047	KEADRGDSGTDLVLEKNKCGKKAAYIKVRVIGSPNSPEGPLEYDDIQRVSRVSWRPPAD	28106	Db	29127	KTYMAKFVKVGTDOVLVKKEISILNIARHRNHLHESFESMEELVMIFESIGLDIFE	29186
QY	426	-----	425	QY	437	-----	436
Db	28107	DGGADILGYLERREVPKAAWYITDSRVRGTSILVVKGLKENVEYHFRVSAENQFGISKPL	28166	Db	29187	RINTSAFELNEREIVSVYHVQCEALQFLSHNHIGHDIRPENIYYOTRRSSTIKIIEFGQ	29246
QY	426	-----	425	QY	437	-----	436
Db	28167	KSEEPVTKTPLNPPPPNPPELVDTKSSVLSWSRPKDDGSRVTGYIIERKETSTD	28226	Db	29247	ARQLKPGDNFRLLFTAPEYAPAEVHQHDVVSTATDMWSLGTLYVVLGSGINPFLAETNOQ	29306
QY	426	-----	425	QY	437	-----	436
Db	28227	KWVRHNKTQITTTMYTWTGLVPDABYQFRIIAQNQVGLSETSPASEPVPVCKPFPDKPSQP	28286	Db	29307	IENIMNAEYTFDEBAFKEISIEAMD FVDRLLVKERKSRMTASEALQHPWLKQKIERVST	29366
QY	426	-----	425	QY	437	-----	436
Db	28287	GELEILSISKDSVTLOWKXPECDDGGKEILGYWVEYRQSGDSAWKKSNERIKDKQFTIGG	28346	Db	29367	KVIRTLKHRRYYHTLIKKDLNMVVSAAIRSCGAIRSQKGVSAKVVASIBIGPVSQOI	29426
QY	426	-----	425	QY	437	-----	436
Db	28347	LLEATEYFRVAENETGLSRPRRTAMSINKTLTSGEAPGIRKEMKDVTTKLGEEAQLSC	28406	Db	29427	MHAVGEGGHVKYVCKIENYDQSTQVTVFQVRQLENSEKEYEITYEDGVAIIYVKDITKL	29486
QY	426	-----	425	QY	437	-----	436
Db	28407	QIVGRPLDPDKWYRFKELIQSRKYKMSDGRHTLTVMTBQDEBGVYTCIATNEVGEV	28466	Db	29487	DDGTYRCKVNDYGEDSSVAELFVKGVREVYDYCRRTWKIKRRTDTMRLLERPPEFTL	29546
QY	426	-----	425	QY	437	-----	436
Db	28467	ETSSKLLQLQATPOFHPGYPLEKXYGAVGSTLRLHWYIGRPVPAMTWFGQKLLQNSN	28526	Db	29547	PLYNKTA YGVNVRFGVTITVHPPEHVTWYKSGQIKPGDNDKKYTFESDKGLYQLTINS	29606
QY	426	-----	425	QY	437	-----	436
Db	28527	ITIENTEHYTHLVKMNQVRKTHAGKYVQLSNVFGTVDAILDVEIQDKPKPTGPIVIEA	28586	Db	29607	VTTDDDAEYTVVARNKYGEDSCAKLTVTILHPPPTDSTLRPMFKRLLANAECQEGSVCF	29666
QY	426	-----	425	QY	437	-----	436
Db	28587	LLKNSAVISWKPADDGGSWITNYVVEKCEKEGAEWQLVSSAISVTTCRIVNLTENAGY	28646	Db	29667	EIRVSGIPPTLKEKGQPLSLGPNIEIIHEGLDYALHIRDITLPEDTGYRVRTATNTA	29726
QY	426	-----	425	QY	437	-----	436
Db	28647	YFRVSAQNTFGISDPLEVSSWIIKSPFEKPGAPGKPTITAVTKDSQVVAWKPPASDGA	28706	Db	29727	GSTSCQALHQLBERLYKQEFKSEHEHVKQIDKTLRMAEILSGTESVPLTQVAKEA	29786
QY	426	-----	425	QY	437	-----	436
Db	28707	KIRNYLEKREKQKWI SVTTEIRETVFSVKNLJIEGLE YEFVRKCNLGESEWSEIS	28766	Db	29787	LREAAVLKPAVSTKTIVKGEFLEIEEKKERKL RMPYDVPEPRKYKQTTIEEDQRIKQF	29846
QY	426	-----FLDEIRNFSLR-----	436	QY	437	-----	436
Db	28767	EPITPKSDVPIQAPHFKBELNLNRYQSNATLVCKVTGHPKPIVKWYRQKEIIADGLK	28826	Db	29847	VPMDSMKWYKIRDOQVEMPGKLD R VVQKRPKRIRLSRWEQLLYQVRPKWRIPKLSQDDLE	29906
QY	437	-----	436	QY	437	-----	436
Db	28827	YRIQEPKGGYHQLIIASVTDDATVYQVRATNOGGSVSGTASLEVEVPAKIHLPKTLEGM	28886	Db	29907	IVRPARRTPSPDYFYVRPRRSLGDISDELLLPIDDDYLAWKRTTEERLRLEEBELG	29966
QY	437	-----	436	QY	437	-----	436
Db	28887	GAVHALRGEVVISIKIPFSGKPDVITWQGDLDINNHYQVIVTRSF TSLVPNGVERK	28946	Db	29967	FSASPPSRPPHPELSSLYSSPQAHVKEETRKOPFRYSTYHIPTKAEASTSYAELRERH	30026
QY	437	-----	436	QY	437	-----	436
Db	28947	DAGFYVVCANRRFGIDQKTVELDVADVPDPRGVKVS DVS RDSVNLTWTEPASDGGSKIT	29006	Db	30027	AQAAYRQPKQRQIRMAEREDELLRPVTTTQHLSEYKSELDFMSKEEKSRRQREVT	30086
QY	437	-----	436	QY	437	-----SPTYVNL-----	443
Db	29007	NYIVEKATTAEWLVRVGQARETRYTVINLFGKTSYQFRVIAENKFGLSKPSBSPTIT	29066	Db	30087	EITEIEEYEIFSKHAQRESSSASRLRRRSLSPSTYIELMRPVSELIRSRQPABEYED	30146
QY	437	-----	436	QY	444	-----TP-----	445
				Db	30147	DTERRSPFTRTPRSPSPVSSRSLSRFRSARPDIFSRYESMKAALKTOKTTSERKYEY	30206

```
QY 446 ----- 445
Db 30207 LSQPFTHDAPRITILMRSHRVPCCGQNTFRILNVQSKPTAEVKWYHNGVELQESSKIHY 30266
QY 446 ----- 445
Db 30267 TWTSGVLTLLEILDCHTDDSGTYRAVCTNYKGEASDYATLDVTGDTTYASQRRDEVRP 30326
QY 446 ----- 445
Db 30327 SVPPELTRTEAVVSFKTSEMEASSVREKVSQMTETRESLSVYEHSAEMKSAALE 30386
QY 446 ----- 445
Db 30387 EKSLEBKSTRKIKTTLAARILTKPRSMVTYEGESARFSCDTGDGPVPTVTLRKQVLS 30446
QY 446 ----- 445
Db 30447 TSARHQVTTTKYSTPEISSVQASDEGNYSVVENSEGKQAEFTLTIOKARVTEKAVTS 30506
QY 446 ----- 452
Db 30507 PRVKSPERVKSPKAVKSPKRVKSPKSPKAVSPKPTPTKPKVQHLVPSAPPKIQ 30566
QY 453 ----- NLS 455
Db 30567 FLKAEASKIATLCTVVESSVLRKEVTWYKDGKCLKENGHFQFYSADGTYELKINNLT 30626
QY 456 TSN----- 458
Db 30627 ESDQGEYVCEISGEGGTSKTNLQFMGQAPKSIHEKYSKISYETKSKDQKTTSTVTRKTP 30686
QY 459 ----- SLW----- 461
Db 30687 KAPEPISSKPVIVTGLQDTTVSSDSVAKPAVKATGEPRPTAIWTKDKAITOGGKYLSE 30746
QY 462 ----- 461
Db 30747 DRGGFPEIHKTDSDGLYCTVKNAGSVSSCKLTIKAIKDTBAQKVSTQKTSEITP 30806
QY 462 ----- 461
Db 30807 QKXAVVQEBISOKALRSEIKMSEAKSQBKALKEEASKVLISEEVKKSAAATSLKSIYH 30866
QY 462 ----- W----- 462
Db 30867 EBITKTSQASEVRTHAEIKAPSTQMSINEGQRLVLKANIAGATDVKKVNLGVELTNSEE 30926
QY 463 ----- 462
Db 30927 YRYGVSGDQTLTIKQASHRDEGILTCISKTEGIVKCOYDLTSLKSLSDAPAFISQPRS 30986
QY 463 ----- WLQ 465
Db 30987 QNINEQNVLFTCEISGEPSEIENFK 31013
```

```
RESULT 4
ADQ17316
ID ADQ17316 standard; protein; 26926 AA.
XX
AC ADQ17316,
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 133.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
OS Homo sapiens.
PN W02004048938-A2.
```

```
XX 10-JUN-2004.
PD
XX 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
DR Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 133; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 26926 AA;
```

```
Query Match 27.8%; Score 681; DB 8; Length 26926;
Best Local Similarity 1.0%; Pred. NO. 1.2e-06;
Matches 268; Conservative 91; Mismatches 106; Indels 25559; Gaps 84;

QY 1 MGRK----- 4
Db 649 VGKKAENVATVAVDQARVREPFGHLESYAQOITLEYGYKERISAQVAPQPPA 708
QY 5 ----- 4
Db 709 SEPHVVPKAVPRVQAPSETHIKTTDDKGMHISQIKKTTDLTTERLVHVDKRPRTASP 768
QY 5 ----- 4
Db 769 HFTVSKISVPKTEHGYEASAGSAIATLQKELSATSSAQITKSVKAPTVPKPSRVAE 828
QY 5 ----- 4
Db 829 PTPLPQFPFADPTDYKSEAGVEVKVEGVSITGTTVREERFEVLHGREAKVTETARVA 888
QY 5 ----- 4
Db 889 PVEIPVPTPLVSLGKNVTVIEGESVTLIECHISGYPSPTVTVREDYQIESIDFQITQ 948
QY 5 ----EMVVRD----- 10
Db 949 SGIARLMIREAPAFEDSGRFTCSAVNEAGTVSTCYLAVQVSEFEKETTAATEKFTTEEK 1008
QY 11 ----- 10
Db 1009 RFVESRDVVMVTDTSLTETQAGGPEAPAFYITKPVVQKLVEGGSVFGQVGNPKPHYV 1068
QY 11 ----- 10
Db 1069 WKSGVPLTTGYRYKVYNKQTGECKLVISMTFADDAGEYTIIVRNKHGSETSASALLEE 1128
QY 11 ----- 10
```

Db	1129	ADYELLMKSQOEWLYQTVTAFOPEVGETAGPFFVSEYEKEQALIRKQWAKDTV	1188	Db	2209	KTTAKLIVEGAVVEFVKELQDIEVPSSYSGELECIIVSPENIEGKWHNDVELKSNKYTI	2268
Qy	11	-----	10	Qy	56	-----	-----DOLK 59
Db	1189	VVRTVDEOEFHISSEERLIKEIEVRIIKTTLEBLEDGEKMAVDISESEAVESGFD	1248	Db	2269	TSRRGRNLTVKQVTKEDQGEYSFVIDGKTKTCKLKMKPRPIALLOGLSQKQCEGDIVQ	2328
Qy	11	-----	13	Qy	60	LEILRQLET	-----TI-----70
Db	1249	LRIKNVRILEGWTFHCKWGYLPFKIAWYKDGKRIKHGERYQMDFLQDGRASLRIPV	1308	Db	2329	LEVKVSLESVEGVMKQGEVQPSDRVHIVIDKQSHMLLIJEDMTKEDAGNYSFTIPALG	2388
Qy	14	-----	13	Qy	71	-----	-----70
Db	1309	LPEDEGITAFASNIKGNACSGKLYVEPAAPLGAFTYIPTLEPVSRIKSLSPRSVSRSP	1368	Db	2389	STSGRVSVSYVDVITPLKDVNVNIEGTAKVLECKVSVDPDTSVKWYLNDEQIKPDDRVQAI	2448
Qy	14	-----	30	Qy	71	-----	-----70
Db	1369	IRMSPARMSPARMSPARMSPOGRLEETDESOLRLYKPVFLKPVSPKLEGANC	1428	Db	2449	VKGTKQRLVINRTHASDEGPKLIVGRVETNCNLSVEKIKIIRGLRDLTCTETQNVVFEV	2508
Qy	31	-----	35	Qy	71	-----	-----77
Db	1429	REFDKVGRMPETFWFHDGQIVNDYTHKVVIKEDGTQSLIIVPATPSDSEWTVVAQN	1488	Db	2509	ELSHSGIDVLNFPKDKBIKPSKYKIEAHGKIYKLTVLNMMKODEGKYTFYAGENMTSGK	2568
Qy	36	-----	35	Qy	78	-----	-----77
Db	1489	RAGRSSIVILTVEAVEHQVPMFVEKLNKVNIEKESRLEMKVNRATGNPNPDIVLKNKSD	1548	Db	2569	LTVAGGAIKPLTDQTVASQEAUFCEVAVNPDSKGEWLDRDKHLPLTNNIRSBDGHR	2628
Qy	36	-----	35	Qy	78	-----	-----86
Db	1549	IIVPHKVPKIRIBGTGEAALKIDSTVQSDSAWYTATANKAGRDTTRCKVNVVEFAEP	1608	Db	2629	RLIIAATKLDIDIGVTVKATSKTAKLKVEAVKIKTKLKNLTVTETQDAVTVVELTHPN	2688
Qy	36	-----	35	Qy	87	-----	-----86
Db	1609	EPERKLIIPRGTYRAKEIAAPELEPLHLRYGQEWEEGDLDYKEKQKQPFKKLTSRL	1668	Db	2689	VKGQWIKNGVLESNEKIAISVKGTIYSUKIRKCAIVDESIVYGRGLRGLSARLHVET	2748
Qy	36	-----	35	Qy	87	-----	-----86
Db	1669	KRGPAPHECRLTPISDPTMVVWELHDGKPLEAANRLMINEFGYCSLDYGVAYSRDSGI	1728	Db	2749	VKIIKPKDVTALENATVAFVSVSHDTPVVKWFHKSVEIKPSDKHRLVSRKVKHKLMLQ	2808
Qy	36	-----	35	Qy	87	-----	-----92
Db	1729	ITCRATNKYGTDHTSATLIVKDEKSLVESQLEPRKGLQRIEELERMAHEGALTGVTTD	1788	Db	2809	NISPSDAGEYTAUVQGLECKAKLFVETLHITKTWKNIIEVPETKTASFECEVSHFNVPMSW	2868
Qy	36	-----	43	Qy	93	-----	-----92
Db	1789	QKEKQKDDIIVLYPEFVRVLEGETARFCRVGYPOPKVWYLNGLQIRKSRFRVYDGI	1848	Db	2869	LKNGVEIEMSEKFKIVQGLHQLIIMTSTEDSABYTFVCGNDOVSATLITPIMITSM	2928
Qy	44	-----	50	Qy	93	-----	-----92
Db	1849	HYLDIVDCKSYDTGEVKVTAENPEGVIEHKVKLEIQOEDFRSVLRRAPRPEPHVHEP	1908	Db	2929	LKDINAEKDTITFEVTVNYEGISYKWLKNGVEIKSTDKCQWRTKKLTHSLNIRNVHFGD	2988
Qy	51	-----	50	Qy	93	-----	-----92
Db	1909	GKLOFEVQKVDPRPVDTTTETKEVVKLKRABRITHEKVPSESEELRSKPKRRTESGYEAIT	1968	Db	2989	AADYTFVAGKATSTATLYVEARHIBFRKHIKDIKVLEKGRAMFECEVSEPDITVQMMKDD	3048
Qy	51	-----	50	Qy	93	-----	-----98
Db	1969	AVELSKRKDBSEYELLARKTKDELLHWTBELTEBEKKALAEKGKITIPTFKPKDIELSPS	2028	Db	3049	QELQITDRIKIOKEKYVHRLIIPSTRMSDAGKYTVVAGNVSTAKLFVEGRDVRIRSICK	3108
Qy	51	-----	50	Qy	99	-----	-----98
Db	2029	MEAPKIPERIQSTVQGSDAHFRVRVVGKPDPECEWYKNGVKIERSDRIYWYWPEDNVC	2088	Db	3109	EVQVIEKQRAVVEFEVNEDDVDHAWYKDGIEINFQVQBRHKYVVERRIHRMFISETROSD	3168
Qy	51	-----	50	Qy	99	-----	-----98
Db	2089	ELVIRDVTAEDSASIMVKAINIAGETSSHAFLLVQAKQLITFTQELQDVVAKEDWTATF	2148	Db	3169	AGEYTFVAGNRSSVTLYVNAPEPPVLOELQPVTVQSGKPARFCAMISGRPOPKISWYK	3228
Qy	51	-----	50	Qy	99	-----	-----98
Db	2149	ECETSEPFVKVYKDGMEVHEGDKYRMHSRDKVHFLSILITDSDAEDYSCVLVEDENV	2208	Db	3229	BEQLSTGPKCKFLHDQGEYTLIIIEAFPEDAAYVTCEAKNDYGVATTASLSVEVPEV	3288
Qy	51	-----	55	Qy	99	-----	-----102
		-----GKYKL 55		Db	3289	SPDQEMPVYPPIITPLQDVTVSEGPARFQCRVSGTDLKVSWSKDKKIKPSRFRMTQ	3348

QY 103 ----- 102  
Db 3349 FEDTYQLEIAEAYPEDEGTYTFVANNAGVQSVSTANLSLEAPESILHERIEQIEMEMKE 3408  
QY 103 ----- 102  
Db 3409 FSSSPLSABEEGHSALQLSKINETLELLSESPVYTPKDPSEKGTGPIFIKEVSNADI 3468  
QY 103 ----- 102  
Db 3469 SMDVATLSVTYIGIPKPIQWFPNGVLLTPSADYKFPVGDGDDHSLIILFTKLEDEGEYT 3528  
QY 103 ----- 102  
Db 3529 CMASNDYGKTI CSAYLKINSKGEGKDTETESAVAKSLEKLGCPHFKELKPIRCAQ 3588  
QY 103 ----- 110  
Db 3589 GLPAIFEYTVGSPAPTVMFKENKQCTSVYVYTIHNPNGSGTFIVDPQREDSGLYIC 3648  
QY 111 ----- 121  
Db 3649 KAENMLGESTCAEALLVLEDDTMDTDPCKAKSTPEAPDPQTPLKGPFAVEALDSEQEI 3708  
QY 122 ----- 121  
Db 3709 ATPVKDTILKAALITEENQQLSYEHIAKANELSSQLPLGAQELQSILODKLTPESTREF 3768  
QY 122 ----- 121  
Db 3769 LCINGSIHFPQPKPSNQLQIVSQKTSKEGILMPEPEPQAVLSDEKIFPSAMSI 3828  
QY 122 ----- 121  
Db 3829 EQINSLTVPLKTLAEPGPNYPOSSIEPPMHSYLTSVAEVLSLKEKTVSDTNREQRTV 3888  
QY 122 ----- 121  
Db 3889 LQKQEAQSALILSQSLAEGHVESLQSPDVMISQVNYEPLVPSHSCTEGGKILIESANPL 3948  
QY 122 ----- 121  
Db 3949 ENAGQDSAVRIERGKSLRFPPLALEEKQVLLKEHSDNVMPDPQIIIESKREPVAIKKQES 4008  
QY 122 ----- 121  
Db 4009 VQGRDLLSKESLLSGIPBEOQLNLKIQICRALQAASVASEQPLFSEWLNRNIEKVEVEAVN 4068  
QY 122 ----- 121  
Db 4069 ITQEPRHMCMLVTSKSVTEBVTIIEVDVDPQMANLKMELRDALCAIIEBIDILTAE 4128  
QY 122 ----- 121  
Db 4129 GPRIOQAKTSLOEEMDSFSGQKVPITEPEVESKYLISTEBVSFNVQSVKYLDATP 4188  
QY 122 ----- 121  
Db 4189 VTKGVASAVVSBKQDESLLKPSSEKESSSESCTEBVATVKIQEABGGLIKEDGPMIHTP 4248  
QY 122 ----- 121  
Db 4249 LVDTVSEEGDIVHLTTSITNAKEVNVYFENKLVPSDEKFKCLQDQNTYTLVIDKVNTEDH 4308  
QY 122 --KYVYSQVNHAKTI- 135  
Db 4309 QGEYVCEALNDSKGTATSAKLTVVKGAAAPVIKRIEPLVALGLAKFTCEIQSAPNRF 4368  
QY 136 ----- 135  
Db 4369 QWPKAGREIYBCKSIRSKYITSISLEILRTQVDDCGEYTCASBYSGVSTATLTVTV 4428

QY 136 ----- 135  
Db 4429 PGGEKKVRKLLPERKPEKEEVVLKSVLRKPEEBEPKVEPKLEKVKKPAVPEPPPKP 4488  
QY 136 ----- 141  
Db 4489 VEEVEVPTTKBERKIPBTKVPEIKPAIPLAPBPKPKPEAEVTKIKPPVPPEBPTPIA 4548  
QY 142 ----- 141  
Db 4549 APVTVPVVGKAEAKAPKEEAAKPGPIKGVPKTSPSEABRRKLRPGSGGKPPDEAP 4608  
QY 142 ----- 141  
Db 4609 FTYQLKAVPLKFVKEIKOIIILTESEBFVSSAIFECILVSPSTAITTTMMKGSNIRESPKHR 4668  
QY 142 ----- 141  
Db 4669 FIADGKDRKHLIIDVQLSDAGEYTCVLRGNKEKSTAKLVVEELPVRPVKTLSEETVV 4728  
QY 142 ----- 145  
Db 4729 KGOPYLSCELNKERDVMWRKDKIWEKPGRIVPGVIGLMRALTINDADDDTAGTYTVT 4788  
QY 146 ----- 145  
Db 4789 VENANNLECCSCVKVVEVIRDMVKPIRDQHVKPKGTAFACDIAKDTPNIKWFKGYDEI 4848  
QY 146 ----- 145  
Db 4849 PAEPNDKTEILRDGNHLYDKIRNAMPEDIAEVAVEIEGKRYPAKUTLGSREVELLKPIED 4908  
QY 146 ----- 150  
Db 4909 VTIIYKESASFDAEISEADI PGQWKLKCELLRSPSTCEIKABSGKRFLLHKVKLQAGE 4968  
QY 151 ----- 150  
Db 4969 VLYQALNAITTAITLTVKEIELDPVPLKDVTPERRQARFECVLTREANVINWISKGPDIK 5028  
QY 151 ----- 150  
Db 5029 SSDKFDIADGKKHILVINDSQFDDGEVYTAEBEGKTSARLFTVGTIRLKFMSPLSDQTV 5088  
QY 151 ----- 150  
Db 5089 KEGETATFVCELSHSHQVHVWFKNDAKLHSTRVLISSSEKTHKLEKVTLLDDISQIKA 5148  
QY 151 ----- 150  
Db 5149 QVKELSSSTAQLKVLBADPYFTVKLHDKTAVEKDEITLKCEVSKDVPVKWFKDGEIIVPSP 5208  
QY 151 ----- 150  
Db 5209 KYSIKADGLRRLIKTKKADLKDKGEYVDCGTDKTKANVTVEARLIEVBKPLYGVFVVG 5268  
QY 151 ----- 150  
Db 5269 ETAHFEIELSEPDVHGQWKLKGOPLTASPDCEIIEGKKHILLHNCQLGTMTEVSFQAA 5328  
QY 151 ----- 150  
Db 5329 NAKSAANLKVKBELPLIFITPLSDVKVFEKDEAKFECEVSREPKTFRWLKGTQBITGDDRF 5388  
QY 151 ----- 150  
Db 5389 ELIKDGTKHSWIKSAAFEDEAKYFEBDKHTSGKLIIEGIRLKLFTPLKDVTAKEKES 5448  
QY 151 ----- 150  
Db 5449 AVPTVELSHDNIRVKWPKNDQRLHTRSVSMQDEGKTHSITPFKDLSDIDTSQIRVEAMGM 5508  
QY 151 ----- 150

Db	5509	SSEAKLTVLEGGPYFTGKLQDYGVEKDEVILQCEISKADAPVKWFKDGKEIKPSKNAVI	5568	Db	6589	VIDVPGVRNLEVTETFDGEVSLAWEEPLTDCGSKIIGYVVERRDIKRTWVLTADRAES	6648
Qy	151	-----	150	Qy	155	-----LNY-----	157
Db	5569	KTDGKRMILILKALKSDIGQYTCGGTDKTSGLDIEDREIKLVRPLHSVEVMEETAR	5628	Db	6649	CEFTVTGLOKGGVEYLFYVSARNRVCTGEPVETDNPVEARSKYDVPGLNVITITDVNRF	6708
Qy	151	-----	150	Qy	158	-----	157
Db	5629	PETEISEDIDHANWKLKGEALLQTPDCEIKBERGKIHSVLHNCRLDQGGVDFQAANVKS	5688	Db	6709	GVSLTWBPEYDGCBAITNYVIELRDKTSIRWDTAMTVRAEDLSATVTVDVGEQYVSFRV	6768
Qy	151	-----	150	Qy	158	-----	157
Db	5689	SAHLRVKPRVIGLLRPLKDVTVTAGETATFDCELSYEDIPVEWYLGKKLBSPDKVPRS	5748	Db	6769	RAQNRIGVGKPSAATPFVKVADPIERPSPVNLTSDDQTSQSSVQLKWEPLKDGSPILG	6828
Qy	151	-----	150	Qy	158	-----	157
Db	5749	EGKVHTILTRDVKLEDAGEVOLTAQDKTHANLHVKEPPVEFTKPLEDQTVBEGATAVLE	5808	Db	6829	YIIERCEEGKDNWIRCNMKLVPELTQVTKLEKGNKYLYRVSABNKAGVSDPSEILGLPT	6888
Qy	151	-----	150	Qy	158	-----	157
Db	5809	CEVSRENKVKFKNGTEILSKKYBIYADGRVRLVIHDCPTEDIKTYTCAKDFKTS	5868	Db	6889	ADDAFVBTMDLSAFKDGLEIVVFNPIITILVPGTGYPRPTATWCFGDKVLETGDRVKMT	6948
Qy	151	-----	150	Qy	158	-----	157
Db	5869	NLNVPPHVEFLRPLTDLQVREKEMARFBCELSRENAKVKWFKDGAEIKKKKYDIISKG	5928	Db	6949	LSAYAEELVISPERSDKGIYTLKLENRVKTSIGBEIDVNVIAFPSAPKELKFGDITKDSVH	7008
Qy	151	-----	150	Qy	158	-----	157
Db	5929	AVRILVINKLLDDEABEYSCVETARTSGMLTVLEBEAVFTKNLANIEVSETDTIKLVE	5988	Db	7009	LTWBPPDDGGSPLTGYVVEKREVSRTKTVKMDFVTDLEFTVPDLVQGKEYLKFVCARN	7068
Qy	151	-----	150	Qy	158	-----	157
Db	5989	VSKPGAELVWYKGBEIIETGRYEILTEGRKRILVIONAHLEAGNVCNCLPSSRTDGKY	6048	Db	7069	KCGPEYAYVDEPNMSTPATVPDPENVKWRDTANSIFLTWDPKNDGGSRIKGYVE	7128
Qy	151	-----	150	Qy	158	-----	157
Db	6049	KVHELAAEFISKPNLEILEGEKAEFVCSISKESFPVQWKRDDKTLESKDKYDVIADGKK	6108	Db	7129	RCPRGSDKWACGEPVAETKMEVTLGEBGKWYAYRVKTLNRQGASKPSRPTTEEIOAVDTQ	7188
Qy	151	-----	150	Qy	158	-----	157
Db	6109	RVLVVKQATLQDMGYVMVWGAARAAHLTVIEKLRIVWPLKDRVKEQEVVFNCEVNT	6168	Db	7189	EAPFIFDLVKLAGLTVKAGTKIELPATVTKPEPKITWTKADMLKQDKRITTIENVPKK	7248
Qy	151	-----	150	Qy	158	-----SKRNDTG-----	164
Db	6169	EGAKAKWFRNEAIFDSSKYIILQKDLVYTLRIRDAHLDDQANYVSLTNHNGENVKSA	6228	Db	7249	SIVTIVDSKSDTGYTIIIEAVNVCGRATAVVEVNVLDKPGPPAAFDITDVTNESCLLTMN	7308
Qy	151	-----	150	Qy	165	-----	164
Db	6229	NLIVBEEDLRIVEPLKDIETMEKKSVTFWCKVNRNLNVLTKWTKNGEVPFNDNRVSYRVDK	6288	Db	7309	PPRDDGSKITNYVVERRATDSEVWHKLSSTVKDTNFKATKLIIPNKEYIFRVAENMYGA	7368
Qy	151	-----	150	Qy	165	-----	164
Db	6289	YKHMLTIKDCGPPDRGEYIVTAGQKSVAELLIIIEAPTEFVEHLEDQTVTEFDDAVFSCQ	6348	Db	7369	GEPVQASPIKAYQDPGPPPTRLPSDITKDAVLTWCEPDDGSGSPITGYVVERLDPD	7428
Qy	151	-----	150	Qy	165	-----	164
Db	6349	LSREKANVWYRNGREIKEGKYKPEKDGSIHRLIIKDCRLDDECEYACGVEDRKSARL	6408	Db	7429	TDKWVRCKMPVKDITYRVKGLTNKKYRFRVLAENLAGPGKPSKSTBILIKDIDPDPW	7488
Qy	151	-----LSEM-----	154	Qy	165	-----	164
Db	6409	FVEEIPVEIIRPPQDILEAPGADVFLAELNKDKVEVQWLNNVVVQGDKHQMSBGI	6468	Db	7489	PPGKFTVKDVGKTSVKLNWTKPEHDGGAKIESYVIEMLKTGTDEWVRVAEGVPTTQHLLP	7548
Qy	155	-----	154	Qy	165	-----	164
Db	6469	HRLOICDIKPROQGEYRPIAKDKAARAKLELAAAPKIKTADQDLVVDVGKPLTMVVPYDA	6528	Db	7549	GLMEGQYSFRVAVNKAAGESEPSDPVLCREKLYPPSPRWLEVINITKNTADLKW	7608
Qy	155	-----	154	Qy	165	-----	164
Db	6529	YPKAEAWFKENEPLSTKTIIDTAEQTSFRILEAKKDGKGRYKIVLQNHKGKAEGFNLK	6588	Db	7609	VPEKDGSGSPITNYIVBKRDRVRKGMQTVDTTVKDKCTVPLTEGSLYVFRVAENAIQ	7668
Qy	155	-----	154	Qy	165	-----	164
Db	7669	SDYTEIEDSVLAKDTFTTGPYPVALAVDVTKRHVDLKWEPKNDGGRPIQRYVIEKSR	7728	Db	7669	SDYTEIEDSVLAKDTFTTGPYPVALAVDVTKRHVDLKWEPKNDGGRPIQRYVIEKSR	7728

Qy	165	-----E0GCG-----	169
Db	7729	LGTRWVAGTAGDCNFRVTDVIEGTEVQVRAENEAGVGHPSPTTEILSTEDTSP	7788
Qy	170	-----	169
Db	7789	SPPLDLHVTDAGRKHIAIAWKPEKGGSPFIYGHVEMCPVTEKWMVRNSRPIKOLKFK	7848
Qy	170	-----	169
Db	7849	VEEGVDPKEYVLRVRAVNAIGVSESEISENVVAKDPDCKPTIDLETHDIIVIEGEKLS	7908
Qy	170	-----	169
Db	7909	IPVFRAPVPVTVSHKDGKEVKASDRLTMKNDHISAHLEVPKSVRADAGIYITLENKL	7968
Qy	170	-----	169
Db	7969	GSATASINVKIGLPGCKDIKASDITKSSCKLTWPEPFDDGTPILHYVLERREAGRRT	8028
Qy	170	-----	169
Db	8029	YIPVMSGENKLSWTKDLIPNGEYFFRVKAVNKGGEYIELKNPVIADPKQPPDPVD	8088
Qy	170	-----	169
Db	8089	VEVHNPTAEAMTITWKPLDYDGGKIMGYIIIEKIAKEERWKRNEHLVPILTYTAKGLE	8148
Qy	170	-----	169
Db	8149	EGKEYQFRVRAENAAGISEPSRATPTTKAVDPIDAPKVILRTSLEVKRGDEIADASISG	8208
Qy	170	-----	169
Db	8209	SPYPTITWIKDENIVPEEIKKRAAPLVRRRKGEVEEPPVLPLOTLSIDNSKKGESQ	8268
Qy	170	-----	169
Db	8269	LRVRDSLRLPDHGLYMIKVENDHGIAPACTVSLDTPGPPINFVEDIRKTSVLCKWEPP	8328
Qy	170	-----NFT-----	172
Db	8329	LDGSGEIIINYLTLEKKDKTKPDSEWIVVSTLHRCKYSVTKLIEGKEYLFRVRAENRFGP	8388
Qy	173	-----TFNPMF-----	178
Db	8389	GPPCVSKPLVAKDPFGPPDAPDKPIVEDVTSNMLVKNWPKONGSPILGYWLEKEVNS	8448
Qy	179	-----	178
Db	8449	THMSRVNKSLLNALKANVDLGLLEGLTVFRVCAENAAAGPKFSPSPDPKTAHDPISPPGP	8508
Qy	179	-----	178
Db	8509	PIPRVTDTSSTTIELEWEPPAFNGGEBIVGYFVDKOLVGTNKSRCRTEKMIKRVYTVKE	8568
Qy	179	-----	178
Db	8569	IREGADYKLRVSAVNAAGEPPGETQPVTVAEPOEPPAVELDVSVKGGIIMAGKTLRIP	8628
Qy	179	-----	178
Db	8629	AVVTRPVPTKWTKEEGELDKDRVIDNVGTKSELIIKDALRKHCHRYVITATNSCGSK	8688
Qy	179	-----	178
Db	8689	FAAAVEVFDVPGVLDLKPVTNRKMLNWSDEDDGGSEITGTIERKDAKMTWRQ	8748
Qy	179	-----	178
Db	8749	PIETERSKCDITGLLEGQYKFRVIAKNKFGCGPPVEIGPILAVDPLGPPTSERLTYTE	8808

Qy	179	-----	178
Db	8809	RORSTITLDWKEPRNSGGSPIOGYIIIEKRHRDKDPFERNKRLCPTTSTFLVENLDEHOMY	8868
Qy	179	-----	178
Db	8869	EFRVKAINEIGSESPSLPLNVVIOQDEVPTIKLRLSVRGDTIKVXAGEPVHIPADVTGL	8928
Qy	179	-----	178
Db	8929	PMPKIEWSKNETVIEKPTDALQITKEEVSERSEAKTELSIPKAVREDKGTVTVTASNRLGS	8988
Qy	179	-----	178
Db	8989	VFRNVHVEVYDRPSPRNLAVTDIKAESCYLTDAPLDNGSGSEITHYVIDKRDASRKAE	9048
Qy	179	-----	178
Db	9049	WEVNTVAVEKRYGIWKLIIPNGQYEFVRVAVNKGISDECKSKVVIQDPYRLPGPPGKP	9108
Qy	179	-----FNV 181 	181
Db	9109	KVLARTKSMLYSWTPTPLDNGGSPITGYWLEKEEGSPYMSRVSRAPITKVGLKGVEFNV	9168
Qy	182	PR-----	183
Db	9169	PRLEGVKTYQFRAMAINAAGIGPPSEPSDPEVAGDPIFPPPPSPCEVKDKTKSSISLW	9228
Qy	184	-----	183
Db	9229	KPPAKDGGSPIKGYIVEMOEBGTDWKNVNEPKLITTCCEVVPNLKELKRYFRKAVN	9288
Qy	184	-----	183
Db	9289	EAGESEPSDTTGBEIPATDIOEBEVEFIDIGAQDCLVCKAGSQIRIPAVIKGREPTPKSSWE	9348
Qy	184	-----	183
Db	9349	FDGAKKAMKGVHDIPEDAQLETAENSSVIIIEPCRSHTGKYSITAKNKAGQKTANCR	9408
Qy	184	-----	183
Db	9409	VKMOVPGPPKOLKVSDITRGCSRLSKWMPDDDDGGDIRIKGYVIEKRTIDGKAWTKVNPDC	9468
Qy	184	-----	183
Db	9469	GSTTFVVPDLLSEQQYFFRVRAENRFGIGPPVETIORTTARDPIYPPDPPIKLKIGLITK	9528
Qy	184	-----	183
Db	9529	NTVHLSWKPKNDGGSPVTHYIVECLAWDPTGTKEAWRQCNRKRDVELOFTVEDLVEGG	9588
Qy	184	-----	183
Db	9589	EYEFVRKAVNAAGVSKPSATVPCDCORPMPSPSIDLKPFMEVEBGTNNVAVAKIGVPF	9648
Qy	184	-----	183
Db	9649	PILTWFKAPPKPDNKEPVLVYDTHVNKLVDVDTCLVIPSQSRSDTGLYITITAVNNLGA	9708
Qy	184	-----	183
Db	9709	SKEMRLNVLRPGPPVGPPIKFESVSADQMTLSWFPKDDGSKIITNVVIEKREANRKTW	9768
Qy	184	-----	183
Db	9769	HVSSEPECTYTIPIKLEGGHEVYFRIMAQNKGIGSEPLDSEPETARNLFSVPGAPKPTV	9828
Qy	184	-----	183
Db	9829	SSVTRNSMTVNWEEPEYDGGSPVTGYWLEMKDOTTSKRWKRVNRDPIKAMTLGVSYKVTGL	9888
Qy	184	-----	183

Db	9889	IEGSDYQFRVYAINAAGVPSLPSDPATARDPIAPPGPPKVTWDTKSSADLEWSPL	9948	Db	10969	ECVVTGLQQGKTYFRVKAENIVGLGLPDTTPIBCEQKLVPPSVELDVKLIIEGLVWKAG	11028
Qy	184	-----	183	Qy	208	-----	207
Db	9949	KDGGSKVTGYIVEYKEGKEWEKDKVTRGTLVVTGLKEGAFYKFRVSAVNIAGIGE	10008	Db	11029	TTVRFPAILIRGVPTAKWTTDGSSEIKTDEHTVETDNFSSVLTIKNCLRRDTGEVQITV	11088
Qy	184	-----	183	Qy	208	-----	207
Db	10009	PGEVTDVIEWKORLNSPDLQLDASVRDRIVVHAGGVIRIIAVVSGKPPPTVWNNERTL	10068	Db	11089	SNAAGSKTVAVHLTVLDVFPPTGPINILDVTPHEMTISWQPPKDDGSGSVINYIVEKOD	11148
Qy	184	-----	183	Qy	208	-----TALLRYAQRNC-----	219
Db	10069	PQETIETTAISSWVIKNQORSHQGVYSLAKNEAGERKKTIIIVDLVDPGVGTPFLA	10128	Db	11149	TRKDTWGVSSGSSKTKLPHLQKCEYFRVRAENKIGVGPPLDSTFTVAKHKESPSPS	11208
Qy	184	-----	183	Qy	220	-----	219
Db	10129	HNLNESCULTWPSPEDDGSPITNVVIEKRESDRANTPVITYTRQNAATVQGLIQGA	10188	Db	11209	PPGKPVVTDITENAAVTSWTLPKSDGSGPITGYMERREVTGKWRVNVKTPIALDKFRVT	11268
Qy	184	-----	191	Qy	220	-----	219
Db	10189	YFFRIAAENSIGMGPFVETSEALVIREPIVTPERPDLVKEVTKNTVLTWNPVKYDGG	10248	Db	11269	GLYEGNTYFRVFAENLAGLSKPSSDPIKACRPDKPPGPPINPKLKOKSRETADLWVT	11328
Qy	192	-----	191	Qy	220	-----	219
Db	10249	SEIINYVLESRLIGTEKFKVKTNDNLLSRKYTVKGLKEGDTYFRVSAVNIQCGKPSFC	10308	Db	11329	KPLSDGSGPILGYVECOQKPGTAGWNRINKDELIRQCAFRVPGLIIEGNEYFRPKAANIV	11388
Qy	192	-----	191	Qy	220	-----	219
Db	10309	TKPITCKDELAPTLHLDPRDKLTIRVGEAPALTGRYSGPKPKVSFWFKDEADVLDDRT	10368	Db	11389	GEGPRELAESVIAKDIHLHPPPEVELDVTCRDVITRVYGOTIRILARVKGRPEPDIWTKE	11448
Qy	192	-----	191	Qy	220	-----	219
Db	10369	HIKTPATLAEKIKAKRSDSGKYCVVVENSTSGKFCQVNVVDHPGPPGVPSFDEVT	10428	Db	11449	GKVLVREKRVLDIQLPRVLEIQIKEAVRADHGKVIISAKNSSGHAQSAIVNVLDRPGPC	11508
Qy	192	-----	191	Qy	220	-----	219
Db	10429	KOYMWISWKPPLDDGSGKITNYIIIEKKEVGKDVWMPVTSASAKTTCKVSKLLEGKDYIFR	10488	Db	11509	QNLKVTWTKENCITISWENPLDNGSGSEITNFIVEYRKPNOKGSIVASDVTKRLIKANLL	11568
Qy	192	-----	191	Qy	220	-----	219
Db	10489	IHAENLYGISDPLVSDSMKADRFRVPDAPQPIVTEVTKDSALVTWKNPHDGGKPTNY	10548	Db	11569	ANNEYFRVCAENKVGVTIETKTPIILAINPIDRPGEPENLHIADKGTFFVLKWRPD	11628
Qy	192	-----	191	Qy	220	-----	219
Db	10549	ILEKETMSKRWARTKDPHPTKFRVPDLLEGQYEFVRVSAENEIGIDSPSPSKPVF	10608	Db	11629	YDGGSPNLVYHVERBLKGSDDWVHKGSIKETHVMDRCVENQIYEFVQTKNEGSED	11688
Qy	192	-----	191	Qy	220	-----	219
Db	10609	AKDPIAKSPFVNPEAIDTTCNSVDLTWQPPRHDGSKILGYIVEYQKVGDEWRRANHT	10668	Db	11689	WVKTEBWWVKEDLOKPVLDKLSGLTVKAGDTIRLEAGVRGKPPPEVAWTKDADTDLT	11748
Qy	192	-----	191	Qy	220	-----	219
Db	10669	PESCPETKYKVTGLRDGQTYKFRVLAVNAGESDPAHVPEPVLVKDRLEPPELILDANMA	10728	Db	11749	RSFRVKIDTRADSSKFSLTAKRSDGGKVVVTTATNTAGSFVAYATVNVLDKPGPVRLKI	11808
Qy	192	-----	198	Qy	220	-----	219
Db	10729	REQHIKVGDTLRLSALIIGVPPPKVTWKEDRADPTKARIDVTPVSGKLEIRNAAHEDGG	10788	Db	11809	VDVSSDRCTVCWDPPEDDGGCEIQNYILEKCTKRMVWSTYSATVLTGTTVTRLIEGNE	11868
Qy	199	-----	198	Qy	220	-----	219
Db	10789	IYSLTVENPAGSKTVSVKVLVDKPGPRDLVSEIRKDSCVLTWKEPDLDDGGSVITNYV	10848	Db	11869	YIFRVAENKIGTPPTESKPVIAKTKYKPGRPDPPEVTKYKSEMTVWNPPEYDGGK	11928
Qy	199	-----	198	Qy	220	-----	219
Db	10849	VERRDVAQWSPLSATSKKSHFAKHLNEGNYLFRVAENQYGRGPPFVETPKPIKALD	10908	Db	11929	SITGYFLEKKEKHSRWPVNKSAIPERRMKVQNLDPDHEYQFRVKAENEIGIGEPSLPS	11988
Qy	199	-----	198	Qy	220	-----	219
Db	10909	PLHPPGPKDLHHVDVDTESLVNWNKPRDGGSPITGYLVEYQEGQDWIKFTVTNL	10968	Db	11989	RPVVAKDPIBPPGPTNFRVVDTKHSITLWGKPVYDGGAPITGYVEMRPKIADASPD	12048
Qy	199	-----	207	Qy	220	-----	219
		-----SQTIVFLGL-----		Db	12049	EGWKRCAAAQVLKREFTVTSLDENQEBYFRVCAQNVQVIGRPAELKEAIKPKELPEPPE	12108



QY	220	-----	219	QY	234	-----	233
Db	12109	IDLDASMRKLVIVRAGCPRLFAIVRGPAPKVTRKVGIDNVVRKQVLDVTWAPLVI	12168	Db	13189	DILIPPEGHADLRKTLILRAGVTWRLYVPVKGRPPPKITWSPKNVNLDRIGLDIKST	13248
QY	220	-----	219	QY	234	-----	233
Db	12169	PNSTRDDSGKSYSLTVNPAGEKAVFVNVRLDTPGVPVSDLKVSDDVTKTSCHVSWAPPEND	12228	Db	13249	DFDTPRCENVNKYDAGKYILILENSCGKKEYTIVVKVLDTPGPPINVTVKEISKDSAYV	13308
QY	220	-----TH-----	227	QY	234	-----	233
Db	12229	GGSQVTHYIVKREADRKTWSTVTPEVKKTSFHVTLNVPNGEYFPRVAVNEVPGVPTD	12288	Db	13309	TWEPPIIDGGSPIINVVQKRDARKSWSTVTTECSKTSFRVNLBEGKSYFPRVFAENE	13368
QY	228	-----	227	QY	234	-----	233
Db	12289	VPKPVLASDPLSPDPKRLKLEATEMTKNSATLAWPLPLRDGGAKIDGYIISYREERQPAD	12348	Db	13369	YIGIDPGETRDVAKASQTPGPVVDLKVRVSVKSSCSIGWKPHSDGSGRIIGYVVDPLTE	13428
QY	228	-----	227	QY	234	-----	233
Db	12349	RWTEYSVVXDLVLVVTGLKEGKYKFRVAARNVAVGSLPREAGVVEAKEQLLPPKILMP	12408	Db	13429	ENKQVRVMSLSLOYSAKDLTSGKEYTFRVSAENENGEGTPTSEITVVARDDVVAPDLK	13488
QY	228	-----	227	QY	234	-----	233
Db	12409	EQITIKAGKLRTEAHVYKPHPTCKKKGEDEVVTSHLAVHKADSSILIIKDVTRKD	12468	Db	13489	GLPDLCLAKENSFRLKIPKGPAPSVSWKGEDPLATDTRVSVSSAVNTTLLIVDC	13548
QY	228	-----	227	QY	234	-----	233
Db	12469	SGYSLTAENSSGTDQKIKVVMADAPPPFPDIDADACSLMHIPLDGGSNIT	12528	Db	13549	QKSDAGKYTITLKNVAGTKEGTISIKVWGKPGIPTGPIKFDEVTAEAMTLKWAPPKDDG	13608
QY	228	-----	227	QY	234	-----	233
Db	12529	NYIVEKDVSRGDWWTALASVTKTSRVGKLIPGQYIFRVAENRFGISEBLTSPQWVA	12588	Db	13609	SEITNYILEKRDVNNKWTVCASAVQKTTFRVTRLHEGMEYTFRVSANIKYGVGEGLKSE	13668
QY	228	-----	227	QY	234	-----	233
Db	12589	QPPFGVPSEPKNARVTKNKDCIFVADWRPDSGSGSPIIGYLLIERKERNLLVVKANDTL	12648	Db	13669	PIVARHPPDVPDAPPPPNIVDVHRDSVSLTWDPKTGGSPITGYHLEFKERNLLWKRA	13728
QY	228	-----	227	QY	234	-----PRV-----	236
Db	12649	VRSTEYPCAGLEGLEYSFRIYALNAKAGSPSPKTEYVTARMPVDPGKPEVIDVTKST	12708	Db	13729	NKTPIMRDFKVTGLTEGLEYEFRVMAINLAGVGRPSLSEFPVALDIPDPGKPEVINI	13788
QY	228	-----	227	QY	237	-----PKYINGTKL-----	245
Db	12709	VSLIWARPKHGGSKIIGYFVEACKLPQDKWVRCNTAPHQIPQEBYTATGLEEKAQYQPR	12768	Db	13789	TRNSVTLLIWEPKYDGGHKLGTGYIVEKRDLPKSKWMAHNVNVPBCAFTVTLVEGGKYE	13848
QY	228	-----	227	QY	246	-----	245
Db	12769	ATARTAVNISPPSESDPVTILAEVPPRIDLSVAMKSLITVKAQTNVCLDATVFGKMP	12828	Db	13849	FRIRAKNTAGASAPSESTETIICKDEYEAPTVILDPTIKDGLTIKAGDTIVLNAISILG	13908
QY	228	-----AMGRNL-----	233	QY	246	-----	245
Db	12829	TVSWKKDGTLLKPAEGIKMAQORNLCTBLEFSVNRKDSGDYTTITAENSSGSKSATIKLV	12888	Db	13909	KPLPKSSWSKAGKDIRPSDITQITSTPTSSMLTIKYATRKDAGEYITITATNPFGTKVEHV	13968
QY	234	-----	233	QY	246	-----	245
Db	12889	LDKPGPPASVKINMYSDRAMLSWEPPLDGGSEITNYIVDKRETSRPNWAQVSATVPIT	12948	Db	13969	KVTVLDVPGPPGPVBEISNVSAEKATLTWTPPLEDGGSPIKSYILEKRETSRLLTWTVVSED	14028
QY	234	-----	233	QY	246	-----K 246	246
Db	12949	SCSVEKLIBGHEYOFRICAENKYGVDPVTEPAKPNYDPPGRCDPPVINSITKDHMT	13008	Db	14029	IQSCRHVATKLIQNGEYIFRVSANVHYKGEPVQSEPVQKQVDRFGPPGPEKPEVSNVTK	14088
QY	234	-----	233	QY	247	NT-----MRKLK-----	253
Db	13009	VSWKPPADDGGSPITGYLLEKRETAQVNNTKVNRKPIIBRTLKATGLQGBTEYEFRTAI	13068	Db	14089	NTATVSWKRPVDDGGSEITGVHVERREKSLRWVRAIKTTPVSDLRCKVTGLQEGSTYFR	14148
QY	234	-----	233	QY	254	-----	253
Db	13069	NKAGPGKPSDASKAAYARDPQPPAPPAPPKVYDTTTRSSVLSWGPAYDGGSPIIGYLV	13128	Db	14149	VSAENRAGIGPPSEASDSVLMDKDAAYPPGPPSNPHVDTTTKKSASLANWKPHYDGGLEIT	14208
QY	234	-----	233	QY	254	-----	253
Db	13129	EVKRADSNWRCNLQNLQKTRFVETGLMEDTQYQFRVAVNKNIGYSDPSDVPDKHYPK	13188	Db	14209	GYVVEHQVGDPAWIKDITGTALRITQFVVVPLQTKKYNFRISAINDAGVGEPAVIPDV	14268
				QY	254	-----RKQAP-----	258

Db	14269	EIVEREMADPFELDAELRLTVVRAGLSIRIFVPIKGRPAPEVTWTKONINLKNRANEN	14328
QY	259	-----SYT-----	281
Db	14329	TESFTLLIPECNRYDTGKFVMTIENPAGKSGFVNVRLDTPGPVLNLRPTDITKDSVT	14388
QY	259	-----	281
Db	14389	LHWDLPIDGSGRIITNYIIVEKREATRKSYSTATTCKCHKTKYKVTGLSEGCYFFRVMAEN	14448
QY	259	-----	281
Db	14449	EYIGICEPTTEBPVKASEAPSPDSLININDITKSVLSLAWPKPKHDGSGKITGYVIEAQR	14508
QY	259	-----VKE-----	261
Db	14509	KGSDQWTHITTVKGLCEVVRNLTEGBEYTFQVMAVNSAGRSAPRESRPVIVKEQTMPEL	14568
QY	262	-----	261
Db	14569	DLRGIYQKLVIAKAGANIKVEIPVLGRPKPTVTWKGDQILKQTORVNEFTTATSTILNI	14628
QY	262	-----	261
Db	14629	NECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPTGPIKDFEVSDFVTFSWDPDEN	14688
QY	262	-----OPEKKA-----	267
Db	14689	DGGVPISNVYVEMRQDSTTWVELATTVIRTTYKATRLTGLLEYQPRVKAQNRYGVGPGI	14748
QY	268	-----	267
Db	14749	TSAMIYANYPFKVPGPPTQVTAVTKDSMTISWHHEPLDGGSPILGYHVERKERNGILW	14808
QY	268	-----	267
Db	14809	QTVSKALVPGNIFKSSGLTDGIAEFVRVIAENMAGSKSPKSEPMALDPIDPPGKVPV	14868
QY	268	-----	267
Db	14869	LNIRTRVTILKWAPEYTGFKITSYIVEKRDLPNGRWLKANFSNILENEFTVSGLTEDA	14928
QY	268	-----	267
Db	14929	AYEFRVIAKNAAGAI SPPSEPSDAITCRDDVEAPKIKVDVKFDTVLKAGEAPRLEADV	14988
QY	268	-----	267
Db	14989	SGRPPTMWSKDGKELEGTALEIKIADFSTNLVNKSDTRDSGAYTLTATNPGGFAXH	15048
QY	268	-----	267
Db	15049	IFNVKVLDRPPEGPEGLAVTEVTSEKCVLSWFPPLDDGGAKIDHIYIVOKRETSRLAWTNV	15108
QY	268	-----	267
Db	15109	ASEVQVTKLVTKLLKGNEXIFRVMVKNYGVGEBLESEPVLA VNPYPGPPDPKQNPVTT	15168
QY	268	-----KKT-----	270
Db	15169	ITKDSMVVCGHPDSGGSEIINIYVERRDRAGQRIWIKCNKKTLDLRYKVSGLTEGHEY	15228
QY	271	-----QSTTTTYPF-----	278
Db	15229	BFRIMAENAGISAPSPSPFYKACDITVFKPGPGPNRVLDTSRSSISIANKKIYDGS	15288
QY	279	-----	278
Db	15289	EITGYMVEIALPEEDEMQIVTPPAGLKATSVTITGLTENQBYKIRIYAMNSEGLGEPALV	15348
QY	279	-----	278
Db	15349	PGTFKABDRMLPBEIELDADLRKVVITIRACCTLRLFPVIKGRPDPPEVKWARDHGESLDKA	15408
QY	279	-----SYT-----	281
Db	15409	SIESASSYTLIIIGNVNRPDGKVIILTVENS S KSAFVNVRLDTPGPQDLKVEVTK	15468
QY	282	-----	281
Db	15469	TSVLTWDPPLLDGSGKIRNYIVEKRESTRKAYSTVATNCHTKSWKVDQLQEGCSYFRV	15528
QY	282	-----	281
Db	15529	LAENEYIGIGLPAETAESVKASERLPPGKITLMDVTNSVLSWEKPEHDGSGSILGYIV	15588
QY	282	-----TSAALNVT-----	289
Db	15589	EMQTKGSKWATCATVKVTEATITGLIQGBEYSPRVSQAONEKGISDRQLSPVPIAKDLV	15648
QY	290	-----	289
Db	15649	IPPAFKLLFNTFTVLAGEDLKVDVPFFIGRPTPAVTWHKDNVPLKQTRVNAESTENNSLL	15708
QY	290	-----	289
Db	15709	TIKDAREDUGHYVVKLTNSAGEAIEITLNVILDKPGPPTGPKVMDEVTA DSITLSWGP	15768
QY	290	-----	289
Db	15769	KYDGGSSINNYIVEKRDSTTTTQIVSATVARTTIKACRLATGCEYQPRIAENRYKGST	15828
QY	290	-----	289
Db	15829	YLNSEPTVAQYFKVPKPGPPGTVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERN	15888
QY	290	-----	289
Db	15889	LWVKLNKTFPIQTKFKTTGLEGEVEYEFVRVSAENIVGIGKPSKVSECYVARDPCDPPGRP	15948
QY	290	-----	289
Db	15949	EALIVTRNSVTLQWKPTYDGGSKITGYIVEKKELPEGRMMKASPTNIIDTHFEVTGLVE	16008
QY	290	-----	289
Db	16009	DHRYEFVRVIA RNAAGVFPSESESTGAITARDEVDPPRISMDPKYKDTIIVWHAGESFKVDA	16068
QY	290	-----	289
Db	16069	DIYGKPIPTIOWIKGQELSNTARLEIKSTDPATSLSVKDAVRVDSGNYILKAKNVAGER	16128
QY	290	-----	289
Db	16129	SVTVNVKVLDRPPEGPEVVISGVTAEKCTLA WKPLQDGGSDIINYIVERRETSRLVMT	16188
QY	290	-----	289
Db	16189	VVDANVQTLSCVKYTKLLEGNEYTFRIMAVNKYGVGEBLESEPVWAKNPVWPDPAPKAPV	16248
QY	290	-----	289
Db	16249	TTVTKDSMIWVERPASDGGSEILGYVLEKRDKEGIRWTRCHKRLIGELRLAVTGLIENH	16308
QY	290	-----	289
Db	16309	DYEFVRVSAENAGLSEPSPSAYQKACDPIYKPGPPNPKVIDITRSSVFLSWSKFIYDG	16368
QY	290	-----	289
Db	16369	GCEIQGYIVEKCDVNVGEWMTCTPPTGINKNTINIEVKLEKHEYNFRI CAINKAGVGEHA	16428
QY	290	-----	289
Db	16429	DVPGPIIVEEKL EAPDIDL DLRLKTIINTRAGSLRLFPVIKGRPTPEVKWGVDEIRD	16488

QY 290 ----- 289  
Db 16489 AAIIDVTSSFTSLVLDNVNRYDSKGYTLTLENSSGTSKSAFVTVRVLDTPSPVNLKVTEI 16548  
QY 290 ----- 289  
Db 16549 TKDSVITWEPPLLDGSGKIKNYIVEKREATRKSYAAVVTNCHNSWKIDQLQBGCSYF 16608  
QY 290 ----- 289  
Db 16609 RVTAENEYGIGLPAQTADPIKVAEVPQPGKIITVDDVTNRNSVLSWTKPEHDGSGKIIQY 16668  
QY 290 -----TNVT----- 293  
Db 16669 IVEMQAKHSEKSECARVKSLOAVITNLTOGBEYLFVRVAVNEKGRDPRSLAVPIVAKD 16728  
QY 294 ----- 293  
Db 16729 LVIEPDVKPAFSYSVQVQDLMKEVPIISGRPKPTITWTKDGLPKQTTRINVTDSLDT 16788  
QY 294 -----YSITTA----- 299  
Db 16789 TLSIKETHKDDGQYGITVANVVQKTASTIEITLDKPPKPVKFDVSAESITLSWN 16848  
QY 300 ----- 299  
Db 16849 PPLYTGCCQITNYIVOKRDTTITVMDVVSATVARTTLKVTCLKTGTGYOFRIPAENRYGQ 16908  
QY 300 ----- 299  
Db 16909 SFALSDPIVAQYPYKEPGPGTPPATAISKDSMVJQMHEPVNNGSPVIGYHLERKERN 16968  
QY 300 ----- 299  
Db 16969 SILWTKVNTKI IHDTOFKAQNLSEGIYEYFRVYAENIVGVKASKNSECYVARDPCDPPG 17028  
QY 300 ----- 299  
Db 17029 TPEPIWKRNEITLQWTKPVYDGGSMITGYIVEKRDLPDGRMKNKFTNNVETQFTVSGL 17088  
QY 300 ----- 299  
Db 17089 TEDQRYEFRVIKNAAGAIKSPSDSTGPIPAKDEVELPRISMDPKPRDTIVNAGETPRL 17148  
QY 300 ----- 299  
Db 17149 EADVHGKPLPTIEWLRGDKIEIESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAG 17208  
QY 300 ----- 299  
Db 17209 SKSFPVNVKVLDRPGPEGPVQVGTSEKCSLTWSPPLQDGGDISHYVVEKRETSRLA 17268  
QY 300 ----- 299  
Db 17269 WTVVASEVNTSLUKTLLEGNEYFRIMAVNKYGVGEPLSAFVLMKNPFLVLPGPSL 17328  
QY 300 -----ARRV----- 303  
Db 17329 EVTNIADSWTCWNRPDSDGGSEIIGYIVEKDRSGIRWIKCNKRITDRLRVRTGLTE 17388  
QY 304 -----STSTIAYR----- 311  
Db 17389 DHEYFRVSAENAAGVGEPSAIVYKACDPVFKPGPPTNAHIVDTTKNSITLAWGKPIY 17448  
QY 312 ----- 311  
Db 17449 DGSSEILGYVVEICKADEEWQIVTQTLRLVTRFBISKLTEHQBYKIRVCALNKVGLGE 17508  
QY 312 ----- 311  
Db 17509 ATSPGTVKPEDKLEAPELDLDELKRGIVVRAGGSARIHIPKGRPMPEITWSREGEF 17568

QY 312 ----- 311  
Db 17569 TDKVOIEKGVNTQLSIDNCDRNDAGKYLKLENSSGSKSAFVTVRVLDTPGPQONLAVK 17628  
QY 312 ----- 311  
Db 17629 EVRKDSAFLWEPPIIDGGAKVKNYVIDKRESTRKAYANVSSKCSKTSFKVENLTEGAIY 17688  
QY 312 ----- 311  
Db 17689 YFRVMAENEFVGVVETVDVAKAAEPPSPGKVTLTDVTSQTSASLWMEKPEHDGGSRLV 17748  
QY 312 ----- 311  
Db 17749 GYVEMQPKGTEKWSIVAESKVCNAVVTGLSSGQEQYQFRVKAINEKGS DPRVLGVVPVIA 17808  
QY 312 ----- 311  
Db 17809 KDLITQPSLKLFPNTYSIOAGEDLKIEIPVIGRPRPNISWVKDGBPLKQTRVNVSETAT 17868  
QY 312 ----- 311  
Db 17869 STVLHIKEGNKDDFGKYTVTATNSAGTATENLSVILEKPGPPVPRPDEVSADFWVIS 17928  
QY 312 ----- 311  
Db 17929 WBPPAYTGCGQISNVIWEKRDTTTTTHMVSATVARTTIKITKLTGTGYQFRIPAENRY 17988  
QY 312 ----- 311  
Db 17989 GKSAPLDSKAVIVQVPFKPGPPGTPFTSISKDQMLVQWHEPVNDGGTKIIGYHLEQKE 18048  
QY 312 ----- 311  
Db 18049 KNSILWVKNLTKPIQDTKFTTGLDSEGYEPKVSANIVGIGKPSKSECFFVARDPCDP 18108  
QY 312 -----PDSSFMKS----- 319  
Db 18109 PGRPEAIVITRNNVTLKWKPAYDGGSKITGYIVEKKOLPDGRMKNKASFTNVLETFVS 18168  
QY 320 ----- 319  
Db 18169 GLVEDQRYEFRVIARNAAGNFSEPSDSSGAI TARDEIDAPNASLDPKYKDVIVVHAGETF 18228  
QY 320 ----- 319  
Db 18229 VLEADIRGKPIPDVVWSKDGKLEBETAARMEIKSTIQKTTLVVKDCIRTDGGYILKLSN 18288  
QY 320 ----- 319  
Db 18289 VGGTKSIPITVKVLDORPGSPGPKVTVTAEBKCYLAMNPPLODGGANISHYIIEKRETS 18348  
QY 320 -----IWA----- 322  
Db 18349 RLSWTOVSTEQALNKYVKLLPGNEYIFRVMAVKNYIGICEPLESGPVTACNPKYKPPGP 18408  
QY 323 -----TOLR----- 326  
Db 18409 STPEVSAITKDSMVVTWARPVDDGGTEIESGYILEKEDKESVWTKCNKKTLDLRLRVTG 18468  
QY 327 ----- 326  
Db 18469 LTEGHSYEFRAAENAGVGEPSVPFYRACDALLYPPGPSNPKVTDTSRSSVSLAWSK 18528  
QY 327 ----- 326  
Db 18529 PIYDGGAPVKYVVEVKEAADEWTTCTPTGLQGQFTVTKLKENTEFNFRICAINSEG 18588  
QY 327 ----- 326  
Db 18589 VGEPATLPGSVVAQERIIEPPEIELDADLRKVVVLRASATLRLPVTIKGRPEPEVKWEKAE 18648  
QY 327 ----- 326

Db	18649	GILTRAOIEVTSFTMLVIDNTRFDSGRYNLTLENNSGSKTAFVNVRLDSPSPVNL	18708	Db	19729	DLRTRAYVDTTDSRTSLATIENANRNDGKYTLTIQNLVLSAASLTLVVKVLDTPGPPNIT	19788
Qy	327	-----	326	Qy	331	-----	330
Db	18709	TIREVKDSVTLSEBPLIDGGAKITNVIKRETRKAYATITNCKTKTTFRIENLOEG	18768	Db	19789	VQDVTKEASVLSDWPENDGGAPVKNYHIEKREASKAWSVTNNCNLSYKVTNLQEGA	19848
Qy	327	-----	326	Qy	331	-----	330
Db	18769	CSYFVRLASNEYIGLPAETTEPVKVEPPLPPGGRVTLVDVTRNTATIKWEKPESDGS	18828	Db	19849	IYFVRSGENEFGVIGIPAEKTEKGVKITEKPSPEKLGVTISKDSVSLTWLKEPHDGSR	19908
Qy	327	-----	326	Qy	331	-----	332
Db	18829	KITGVVEMQTSKSEKSTCTQVKTLEATISGLTAGEEYVFRVAAVNEKGRSDPRLQVP	18888	Db	19909	IVHYVVEALEKGQKNWVKCAVAKSTHVHVSGLRENSYFVRFAENQAGLSDPRELLPV	19968
Qy	327	-----	326	Qy	333	-----	332
Db	18889	VIARDIEIKPSVELPFHTFNVKARQOLKIDVPFKGRPOATVNRKDGQTLKETTRNVSS	18948	Db	19969	LIKEQLEPEIDMKNFPHSHVTVVRAGSNLKVDIPIISGKPLPKVTLSDRGVPLKATMRFT	20028
Qy	327	-----	326	Qy	333	-----	332
Db	18949	SKVTLSLSIKEASKEDVGYELCVSNSAGSITVPIITIIVLDRPGPPGPIRIDEVSCDSIT	19008	Db	20029	EITAENLTINLKESVTADAGRYEITAANSSTGTTKAFINIVLDRPGPPGPPVISDITEE	20088
Qy	327	-----	326	Qy	333	-----	332
Db	19009	ISMNPPEYDGGCQISNYIVEKETTSTWHIVSQAVARTSIKIVRLTTGSEYQFRVCAEN	19068	Db	20089	SVTLKWEPPKVDGGSQVTVNYILLKRETSATVWTEVSATVARTMMKVMKLTGTGEEYQFRIK	20148
Qy	327	-----	326	Qy	333	-----	335
Db	19069	RYGKSSYSSAVVAEYFPFPPGPGTPKVHATKSTMLVTQVPVNDGGSRVIGYHLEY	19128	Db	20149	AENRFGISDHDSACVTVKLPYTTTPGPPSTPWTVNTVRESITVGWHEPVSNGSASVGVH	20208
Qy	327	-----	326	Qy	336	-----	335
Db	19129	KERSILWSKANKILLIADTVKVSGLDEGLMYEYRVVAENIAGIGKSKCEVPAPDPC	19188	Db	20209	LEMKDRNSILMQANKLVIRTHFKVTTISAGLIYEFVRVAENAAAGVKPSHPSEPLAI	20268
Qy	327	-----	326	Qy	336	-----	335
Db	19189	PPGQPEVNTNRKSVLSKSPHYDGGAKITGYIVERRELPGDRWLKCNVTNIQETYE	19248	Db	20269	DACEPPNRVTRTIDISKNSVLSWQAPAFDGGSKITGYIVERRDLPGRWTKASFTNVTET	20328
Qy	327	-----	326	Qy	336	-----	339
Db	19249	VTELTEDQRYEFVRVARNADSVSPSESTGPIIVKODVEPPRVMDVFRDVIIVKAGE	19308	Db	20329	QFTISGLTQNSQYEFVRVARNAGSISNPSEVGPITCIDSYGPGVIDLPLEYEVVKYR	20388
Qy	327	-----	326	Qy	340	-----	339
Db	19309	VLKINADIAGRPLPVISWAKDGIETEARTEIISTDNHLLITVKDCIRRDGTQVYVTLK	19368	Db	20389	AGTSVKLRAGISGKPAPTIEMWKDKLOTNALVCVENTTDLASILKADRLNSGCVEL	20448
Qy	327	-----	326	Qy	340	-----	339
Db	19369	NVAGTRSVAVCKVLDKPGPPAGPLEINGLTAEKCSLSWRPQEDGGADIDYHKKRET	19428	Db	20449	KLRNAMASASATIRVOIILDKPGPGPIEFKTVTAEKITLLWRPPADGGAKITHYIVEK	20508
Qy	327	-----	326	Qy	340	-----	339
Db	19429	SHLAWTICBGLQMTSCKVTLLKGNEXIFRVTGNKYGVGEPLSVAKALDFTVPSP	19488	Db	20509	RETSRVVMSVSEHLEECITTTTKIKNEXIFVRVAVNKYIGIPELSDSVVAKNAPVT	20568
Qy	327	-----	326	Qy	340	-----	339
Db	19489	PTSLEITSVTKESNTLCSRPSDGGSEISGYIIEERKNSLRVVRNKKPVYDLRVKST	19548	Db	20569	PGPPGIPVTKITKNSMTVWRSRPIADGSDISGYFLEKRDKSLGWFKVLEKTIERTQ	20628
Qy	327	-----	326	Qy	340	-----	347
Db	19549	GLREGCEYRVAENAAAGLSLPSPLIRABDPVFLSPSPSKPIVDSGKTTITIAWV	19608	Db	20629	KVTGLTNSDYQYRVCAVNAAGQFPSEFSEFYKAADPIDPPGPPAKIRIADSTKSSITL	20688
Qy	327	-----	326	Qy	348	-----	347
Db	19609	KPLFDGAPITGYTVEYKSDDDTKTSIQSLAGTEYTTISGLTTGAEYVFRVSKVNKVA	19668	Db	20689	GWSKPVYDGSAGTVGYVEIROGEBEETVSTKGEVTRTEYVVSNLKPGVNYFRVSAV	20748
Qy	327	-----	326	Qy	348	-----	347
Db	19669	SDPSDSDPQIAKEREPEFLDIDSEMRKTLIVKAGASFTMTVPFRGRVPVNLWSKPD	19728	Db	20749	NCAGQGEPIENPEVQAKDILEAPEIDLVALRTSVIAKAGEDVQVLIIPFKGRPPPTVTW	20808
Qy	327	-----	330	Qy	348	-----	351
				Qy	348	-----	351
				Db	20809	RKDEKNLGS DARYSIENTDSSLLITIPQVTRNDTGKYLTIENGVEPKSSTVSVKVLDT	20868

QY	352	-----	351	QY	364	-----	363
Db	20869	PAACQKLVKHSRGVTWILLMDPLIDGGSPINYYVIEKRDKRTWTSVVSHKCSSTFK	20928	Db	21949	LDTPGPGPITFKDVTGRSATLMDAPLLDGGARIHYYVVEKEASRRSQVISEKCTRQ	22008
QY	352	-----	351	QY	364	-----	363
Db	20929	LIDLSEKTPFFRVLAEBIGEGPCETTEPVKAAEVPAPIRDLMSKDKTSVILSWTK	20988	Db	22009	IFKVNDLAEGVPYYFRVSAVNEYGVEPEPIVATQPPRRDLVVVDTSKSSAVLA	22068
QY	352	-----	351	QY	364	-----	363
Db	20989	PDPDGGSVITEYVVERKKGQWTSAGISKTCEIEVSQLEKQSVLEFPRVFAKNEKGLSD	21048	Db	22069	WLKPDHGGSRITGYLLEMRQKSDLWVEAGHTKQLTFTVERLVEKTEYEFVRVAKNDAG	22128
QY	352	-----	351	QY	364	-----	363
Db	21049	PVTIGPITVKELIITPEVDLSDIPGAQVTVRIGHNVHLELPYKGPSPISMLKQGLPLK	21108	Db	22129	YGEPREAFSSVIEKEPQIEPTADLTGITTQLITCKAGSPFTIDVPISGRPAPKPVTKLEE	22188
QY	352	-----	351	QY	364	-----	363
Db	21109	ESBFVRFSKTENKITLSIKNAKKEHGGKVTVILDNAVCRIAVPIVTITLGPSPKPGPIR	21168	Db	22189	MRLKETDRVSITTTKDRITTLTVKDSMRGSGRYPLTLENTAGVKTPSVTVVIGRPGVT	22248
QY	352	-----	351	QY	364	-----	363
Db	21169	FDEIKADSVILSHVDPNEDNGGEITCYIEKRETSQTNKMWCVSSVARTTFKVPNLVKDA	21228	Db	22249	GPIEVSSVSAESCVLWGEPKDGGTEITNYIVEKRESGTTAQVLNSSVKRTQIKVTHL	22308
QY	352	-----	351	QY	364	-----	370
Db	21229	EYQFVRAENRYGVSQLVSSIIIVAKHQFRIIPGPGPKVIYNTSDGMSLTWDAPVYDGG	21288	Db	22309	TKYMEYSFRVSSSENFRGVSKPLESAPIIAEHPFVPPSAPTRPEVYHVHSANAMSIRWEPY	22368
QY	352	-----	351	QY	371	-----	370
Db	21289	SEVTGFHVEKKERNILMKQVNTSPISGREYRATGLVGLDYQFVYAENSAGLSSPSDP	21348	Db	22369	HDGSGKIIGYVWEKKERNITILWKENKVPCLBCNYKVTGLVEGLEVOFRTYALNAAGVSK	22428
QY	352	-----	351	QY	371	-----	370
Db	21349	SKETLAVSPVDPGTPDYIDVTRITILKNPPLRDGSGKIVGYIEKQGNRWRCNP	21408	Db	22429	ASEASRPIMAQNVPDAPGRPEVTDVTRSTVSLWSAPAYDGGSKVGVYIERKPVSEVD	22488
QY	352	-----	351	QY	371	-----	370
Db	21409	TDVSEQYTVTGLSPGDRYEFRIIARNAVGTISPPSQSGIIMTRDENVPPIVEFGEYF	21468	Db	22489	GRWLKCNVTIVSDNFTVTALSEGDTYEPRLAKNAAGVISKSESTGPVTCRDEYAPPK	22548
QY	352	-----	351	QY	371	-----	370
Db	21469	DGLIIKGESLRKALVQGRPVPRVTWPKDGEIEKRMNMBITNVLGSTSLFVRDATRDH	21528	Db	22549	AELDARLHGLDVTIRAGSDLVLDAAVGGKPEPKIITWKGDKDLCEKVSQVTKGRATA	22608
QY	352	-----	351	QY	371	-----	370
Db	21529	RGVYTVKXNAGSAAKAEIKVKQDTPGVGPPIRFTNITGEKQWTLWDAPLNDGCAPIT	21588	Db	22609	VIKFCORSDSGKYTLTVKNASGTVKAVSVNVKLDSPGCGKLTVSRVTQEKCTLAWSLPQ	22668
QY	352	-----	356	QY	371	-----	370
Db	21589	HYIEKRETSRLAWALIEDKCEAQSYTAIKLINGNEYQFRVSAVNFQVGRPLDSDPVVA	21648	Db	22669	EDGGAEITHYIVERRETSRLNWVIVEGECPTLSYVVVTRLIKNNEYIFRVRVAVNKYGGVP	22728
QY	357	-----	356	QY	371	-----	370
Db	21649	QIQYTVDPAGIPEPSNITGNSITLTWARPSDGGSEIOQYILERREKSTRWVKVISKR	21708	Db	22729	VESEPIVARNSTIPSPGPIPEEVGTGKEHIIIQWTKPSDGGNEISNYLVDKREKESLR	22788
QY	357	-----	356	QY	371	-----	370
Db	21709	PISETRFKVTGLTEGNEYEFHWAENAGVGPASGISRLIKCREPNPPGPTVVKVTDT	21768	Db	22789	WTRVNKYVVYDTRLKVTSLMEGCDYQFRVTAVNAGNSEPSESNFISCREPSYTPGPP	22848
QY	357	-----	356	QY	371	-----	370
Db	21769	SKTTVSLEWSPKVPDGMELIIGYIEBMCKTDLGDHVKNAEACVKTRYTVTDLQAGEEYK	21828	Db	22849	SAPRVVDTTKHSISLAWTKPMYDGGTDIVGYVLEMQEKDQWVHTNATINTEFTVP	22908
QY	357	-----	363	QY	371	-----	370
Db	21829	FRVSAINGAGKGDSCSVTGTIKAVDRLTAPELDIDANFKQTHVVRAGASIRLFIAYOGRP	21888	Db	22909	DLKMGQKYSFRVAANVKGMSYSSIAIEIPERIEIPDLLEADDLKKTVTIRAGASLR	22968
QY	364	-----	363	QY	371	-----	370
Db	21889	TPTAVWSKPDNSLSLRADIIHTDTSFSTLTIVENCNRNDAGKYTLTVENNSGSKSITPTVKV	21948	Db	22969	LMVSVSGRPPPVITWTSKQIDILASRAIDTTESYSLLIYVDKNRYDAGKYTIEAENQSGK	23028
QY				QY	371	-----	370

Db	23029	KSATVLKVDYDTPGPCPSVKVKEVSRDSTVIITWEIPTIDGGAPINNYIVEKREAAARAFK	23088	Db	24109	IATNEVEGETSSKLLIQAATPQFHPGYPLEKYYGAVGSTLRHLHVMYIGRPVPANTWFGH	24168
QY	371	-----TIVG-----	374	QY	410	-----	409
Db	23089	TVTTKSKTLYRISGLVEGTMHYFVRLPENIYIGCEPCTSDAVLSEVPLPAKLEVVD	23148	Db	24169	QKLLQNSENITIENTEHYTHLVKMNQVRKTHAGKYKVQLSNVFGTVDAILLDEIQDKPDK	24228
QY	375	-----	374	QY	410	-----	409
Db	23149	VTKSTVTLAWKPLYDGGSLTGYLEACKAGTERMKVVTLKPTVLEHTVTSLNEGEQY	23208	Db	24229	PTGPVIEALLKNSAVISWKPADDGGSWITNYYVVEKCEAKEGAEWQLVSSAISVTTCRI	24288
QY	375	-----TLDMSL-----	381	QY	410	-----ORTF--IDPL-----	417
Db	23209	LFRIRAQNEKGVSEPRETVTAVTVQDLRLPTIDLSTMPQKTHVPAGRPVELVIPIAGR	23268	Db	24289	VNLTENAGYYPRVAQNTFGISDFLEVSSVVIKSPPEKPGAPGKPTTATVKDSCVVAV	24348
QY	382	-----	381	QY	418	-----	417
Db	23269	PPPAASWFFAGSKLRESERVVETHTKVAKLTIRETTIRDTGTYLELKNVGTGTTSETIK	23328	Db	24349	KPPASDGGAKIRNYLYLEKREKKQNKWISVTTEIRETVFSVKNLIEGLEBYFRVKCNELG	24408
QY	382	-----	381	QY	418	-----	417
Db	23329	VIILKPGPTGPIKIDEIDATISITISWEPPELDGGAPLSGYVVEQORDAHRPGWLVPSES	23388	Db	24409	GESEWSEISEPITPKSDVPIQAPHFKEELRNLRNVYQSNATLVCKVTGHPKPIVKWYRQG	24468
QY	382	-----	381	QY	418	-----	417
Db	23389	VTRSTFKFTLREGNEVPRVAATNRFGIGSYLQSEVIECRSSIRIPGPPETLIQIPDVR	23448	Db	24469	KEIIADGLKYRIOEPKGGHQLIIASVTDDBATVYQVRATNQGGSVSGTASLEVEVPKI	24528
QY	382	-----	381	QY	418	-----	417
Db	23449	DGMTLTWYPEDDGSQVTGYIVERKEVRADRWVRNVKVPVTWRYRSTGLTEGLEYEHR	23508	Db	24529	HLPKTEGMAVHALRGEVWSIKIPFGKPDVITWQGDLLDNNGHYQVIVTRSFSL	24588
QY	382	-----	381	QY	418	-----	417
Db	23509	VTAINARGSKPRSPKPIVAMDPIAPPKQPONPRVTDITRTSVSLAWSPEDEGSKVT	23568	Db	24589	VFPNGVERKAGFYVVCANRFGIDQKTVELDVADVPDPRGKVSASRSVNLWTWEP	24648
QY	382	-----	381	QY	418	-----	417
Db	23569	GYLEMQVDQHEWTKNTTPTKIREYTLTHLPQGAERPRVLACNAGGPGEPAEVPGTV	23628	Db	24649	ASDGGSKITNYIVEKATTAERLWVGQARETRYVINLFGKTSYQFRVIAENKFLSKP	24708
QY	382	-----YNETMFV-----	389	QY	418	-----	417
Db	23629	KVTEMLEYDPYELDERYQEGIFVRQGGVIRLTPIKGGPPICKWYKEGQDISKRAMIAT	23688	Db	24709	SEPSBPTITKEDKTRAMNVEDETVESVMTKASHSTKELYEKYMAIEDLGRGEFGIV	24768
QY	390	-----ENK-----	392	QY	418	-----	417
Db	23689	SETHTELVIKEADGDSGTVDLVLENKCGKAVYIKVRVIGSPNSPEGPLEYDDIQVRSV	23748	Db	24769	HRCVETSKKTYMAKPVKVKGTDQVLVKEISILNIARHNILHLHESFESMEELVMIPE	24828
QY	393	-----	392	QY	418	-----	417
Db	23749	RVSWRPADGGADILGYILERREVPKAAWYTIDSRVGTSLVVKGLKENVEYHFRVSAE	23808	Db	24829	FISGLDIFERINTSAFELNEREIVSVYHQVCEALQFLHSHNIGHFDIRPENIYYQTRRSS	24888
QY	393	-----	392	QY	418	-----	418
Db	23809	NQGISKPLKSEEPVTPKTLNPPPEPPNPPEVLVDTKSSVSLSWRPPKDDGGSRVTGY	23868	Db	24889	TIKIIIFGOARQLKPDGNFRLIFTAPYYAVEVHQHDVVSTATDMMSLGTLVYLLSGIN	24948
QY	393	-----TASDS--NKT-----	406	QY	419	-----DYLSLL-----	425
Db	23869	IERKETSTDKVVRHNKTQIITWTYVTVGLVDAEYQFRIIAQNDVGLSETSPASEPVWCK	23928	Db	24949	PFLAETNQOIIENINMAEYTFDEBAFKEISIEAMDFVDELLVKEKRSMTASEALQHPWL	25008
QY	407	-----MGF-----	409	QY	426	-----	425
Db	23929	DPFDKPSQPGFEILSISKDSVTLQWEKPECDGCKEILGYWVEYRQSGDSAMKSNKERI	23988	Db	25009	KQKIERVSTKVIRTLKHRRYYHTLIKDLNMVVSNAARISCGGAIRSQKGVSAKVKVASI	25068
QY	410	-----	409	QY	426	-----FLDEIRNF-----	433
Db	23989	KDKOFTIGLLEATEYEFVRPAENETGLSRPRRTAMSIKTLTSGEAPGIRKEMKDVTK	24048	Db	25069	EIGPVSQIMHAVGEGGHVKYVKIENYDQSTQVTVWYFGVROLNSEKEYEITYEDGVAI	25128
QY	410	-----	409	QY	434	-----	433
Db	24049	LGEAAQLSCQIVGRPLDKWYRFGKELIQSRKYMSSDGRTHTLVTWTEBEQDEGVYTC	24108	Db	25129	LYVKDITKLDGDTYRCVKVNDYGEDSSYAEELFKVGREVVYDYCYCRRTMKKIKERTDTMRL	25188
QY	410	-----	409	QY	434	-----	433
Db				Db	25189	LERPPFTLPLYNKTAAYGVNVRFGVTITVHPHPHTWYKSGQIKPGDNDKKYTFESDK	25248

QY 434 ----- 433  
Db 25249 GLYQLTINSVTTDDAEYTVVARNKYGEDSCAKLTVTLHPPTDSTLRPMFKRLLANAE 25308  
QY 434 ----- 433  
Db 25309 COEGQVCFEIRVSGIPPTLKWEKDGQPLSLGPNIBIIHEGLDYVALHIRTLPEDTGY 25369  
QY 434 ----- 433  
Db 25369 YRVTAINTAGSTSCQAHQVERLYKKQBFKSEHERHVQKIDKTLRMAELLSGESV 25428  
QY 434 ----- 433  
Db 25429 PLTVAKEALREAAVLKPAVSTKTVKGFBLEIEBKSEERKLMPDVPPEPKYKQTTI 25488  
QY 434 ----- 433  
Db 25489 BEDQRIKQFVPMDSMKYKIRDQYEMPGKLDREVQKPRIRLSRWEQFYVYMLPRITD 25548  
QY 434 ----- 433  
Db 25549 QYRPKWRIPKLSQDDLEIVRPARRTPSPDYFYRPRRSGLDISDELLLPIDDYLAN 25608  
QY 434 ----- SLR----- 436  
Db 25609 KRTEERLRLEBELELGFSPSPSPHPELSSLAYSSPQAHVKVEETKKNFRYSTYHI 25668  
QY 437 ----- 436  
Db 25669 PTKAEASTYAEELRERHAQYRQPKQRORIMAREDEBLLRPVTTQHLSEYKSELDFM 25728  
QY 437 ----- SPTYNL----- 443  
Db 25729 SKEEKRKSRQREVTEITEEEVEISKHAQRESSASALLRRRSLSPYIELMPP 25788  
QY 444 ----- TP----- 445  
Db 25789 VSELIRSRQPAEYEDDTERRSPTPTERTPRSPSPVSSERSLSPERSARFDIPSRYES 25848  
QY 446 ----- 445  
Db 25849 MKAALKTKTSERKVELSQPFTLDHAPRITLRMRSHRVPCCQNTFRILNVQSKPTAEV 25908  
QY 446 ----- 445  
Db 25909 KWHNGVELOESSKIHYTNTSGVLTLLEILDCHTDDSGTYRAVCTNYKGEASDYATLIDVTG 25968  
QY 446 ----- 445  
Db 25969 GDYTTVASQRRDEEVPSPVPPELTRTEAYAVPSFKTSEMEASSSVREKVSQMTETRESL 26028  
QY 446 ----- 445  
Db 26029 SSYEHSASAEKMSAALEESLSEKSTTRIKITTLAARILTKPRSMTYRGESEARFSCDPT 26088  
QY 446 ----- 445  
Db 26089 GEPVPTVTLRKQVLSGARHQVTTTKYSTFEISSVQASDEGNYSVVVENSEGKQAE 26148  
QY 446 ----- PEHRAV----- 452  
Db 26149 FTLTIQARVTEKAVTSPRVKSPPEVKSPPEAVKSPKPKVSPESHKAVSFTETKPTP 26208  
QY 453 ----- 452  
Db 26209 REKVQHPVSPAPPKITQFLKABASKEIAKLTCVWESSVLRAKEVTVYKDGKCLKENGHFQ 26268  
QY 453 ----- NLSTN----- 458  
Db 26269 PHYSADGTVELKINNLTESDQGEYVCEISGEGGTSKTNLQFMGQAFKSIHEKVKISETK 26328

QY 459 ----- SLW 461  
Db 26329 KSDQKTTESTVTRKTEPKAPEISSKPVIVTGLQDTTVSSDSVAKFAVKATGSPRTAIW 26388  
QY 462 ----- 461  
Db 26389 TKDGKAITOGGKYKLSKEDKGGPFLEIHKHTDSDGLTYCTTVKNSAGSVSSCKLTJKAIK 26448  
QY 462 ----- 461  
Db 26449 DTEAQKVSTQKTSSEITPQKAVVQBEISQKALRSBEIKMSEAKSQEKALKKEASKVLIS 26508  
QY 462 ----- 461  
Db 26509 BEVKSAATSLEKSIHVHEITKTSQASEEVRTHAEIKAFSTQMSINEGQRLVLKANIAGA 26568  
QY 462 ----- W----- 462  
Db 26569 TDVKWLVNGVELTNSSEYRYGVSGSDQTLTIKQASHRDEGILTCISKTKEGIVKCYDLT 26628  
QY 463 ----- WLQ 465  
Db 26629 LSKELSDAPAFISQPRSQNINEQNVLFCTCEISGEPSPLEWFK 26672  
RESULT 5  
AAU05396  
ID AAU05396 standard; protein; 26926 AA.  
XX AC AAU05396;  
XX DT 24-OCT-2001 (first entry)  
XX Human titin (connectin) protein sequence.  
XX DE Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;  
XX KW titin-related disease; zebrafish; heart failure; heart disease.  
XX OS Homo sapiens.  
XX PN WO200151666-A1.  
XX PD 19-JUL-2001.  
XX PF 12-JAN-2001; 2001WO-US001212.  
XX PR 12-JAN-2000; 2000US-0175787P.  
XX PA (GEO ) GEN HOSPITAL CORP.  
XX PI Fishman MC;  
XX DR WPI; 2001-451869/48.  
XX DR N-PSDB; AA0505390.  
XX PT Determining if a subject has or is at risk of developing a titin-related  
XX PT disease or condition, particularly heart failures, comprises detecting  
XX PT the presence of a mutation in the titin gene.  
XX PS Disclosure; Page 57-111; 114pp; English.  
XX The present sequence representing human titin (also known as connectin)  
CC is described in an invention relating to a novel method for determining  
CC whether a subject has or is at risk of developing a titin-related disease  
CC or condition. The method comprises analysing a nucleic acid of a sample  
CC from the subject and detecting the presence of a mutation (e.g. the  
CC pickwick mutation in the cardiac specific exon N2B) in the titin gene,  
CC which indicates that the subject has or is at risk of developing a titin-  
CC related disease. The zebrafish which has a phenotype similar to mammalian  
CC heart failure is used as a model. The method is useful for detecting an  
CC increased likelihood of heart disease, such as heart failure, in a  
CC patient, so that appropriate intervention can be instituted before any  
CC symptoms occur. The method may also be used to facilitate determination

CC of etiology of an existing heart condition, such as heart failure, to  
CC identify compounds that can be used to treat or prevent heart conditions,  
CC in prenatal genetic screening, e.g. to identify parents who may be  
CC carriers of a recessive titin mutation. Compounds identified using the  
CC methods may be used to treat patients that have or are at risk of  
CC developing heart disease, e.g. heart failure  
XX  
SQ Sequence 26926 AA;  
  
Query Match 27.7%; Score 678; DB 4; Length 26926;  
Best Local Similarity 1.0%; Pred. No. 1.5e-06;  
Matches 267; Conservative 93; Mismatches 105; Indels 2559; Gaps 85;  
  
QY 1 MGRK----- 4  
DB 649 VGKAEAVATVVAVDQARVREPPECHLEESYAQQTTLLEYGYKERISAQVAEPQRA 708  
QY 5 ----- 4  
DB 709 SEPHVVPKAVPRVIQAPSETHIKTTDQGMHISSQIKKTTDLTTLRLVHVKRPRTASP 768  
QY 5 ----- 4  
DB 769 HFTVSKISVPKTEHGYEASIAIATLOKELSATSSAQKITKSVKAPTVPKSETRVRAE 828  
QY 5 ----- 4  
DB 829 PTPLOPPFADPTDITYKSEAGVEVKVGVISITGTTVREERFVLHGREAKVETARVPA 888  
QY 5 ----- 4  
DB 889 PVEIPVTPPLVSGLKNVTIEGESVTLECHISGYPSTVWYREDYQIESSIDFQITFQ 948  
QY 5 EMMVRD----- 10  
DB 949 SGIALRMIREAFADSGRFTCSAVNEAGTVSTCYLAVQVSEFEKETTAATEKFTTEEK 1008  
QY 11 ----- 10  
DB 1009 RFVESRDVMTDSLTEEQAGPEPAAPYFITKPVVQKLVEGGSVFGCVGQGNPKPHVY 1068  
QY 11 ----- 10  
DB 1069 WKKSGLVPLTGYRYKVSYNKQTGECKLIVISMTFADDAGEYTIIVVRNKHGTSASALLEE 1128  
QY 11 ----- 10  
DB 1129 ADYELLMSQOEMLYQTQVTAFOEPEVGETAPGFVYSEYEKEQEALIRKQAKDTV 1188  
QY 11 ----- 10  
DB 1189 VVRTYVEDQEFHISSPEERLIKIEYRIIKTTLEELLEEDGEKMAVDISESEAVESGFD 1248  
QY 11 VPK----- 13  
DB 1249 LRIKNVRILEGMVTFHCKMGYPPLKIANWKDKRIKHGERYQMDFLQDGRASLRIPVV 1308  
QY 14 ----- 13  
DB 1309 LPEDEGIYTFAPASNIKNAICSGKLYVEPAAPLGAPTYIPTLEPVSRIKSLSPRSVSRSP 1368  
QY 14 ----- 22  
DB 1369 IRMSPARMSPARMSPARMSPGRRLEETDESQLERLYKPVFVLKPVSFKLEQOTA 1428  
QY 23 ----- 22  
DB 1429 RFDLKVVGRMPETFWFHGQQIVNDYTHKVIKEDGTQSLIIVPATPDSGSGTWVVAQN 1488  
QY 23 ----- 22  
DB 1489 RAGRSSIVILTVEAVEHVKMPFVEKLNKVNKEGSRLEMKVATGNPNPDIVLWLNKSD 1548

QY 23 ----- 22  
DB 1549 IIVPHKYPKIRIEGTGGEALKIDSTVSQDSAWYATATAINKAGRDITTRCKVNVVEFAEP 1608  
QY 23 ----LLV----- 25  
DB 1609 EPERKLIIPRGTRYAKEIAAPELPLHLRYQOEOWEGDLYDKEKQKQPPFKKLTSLRL 1668  
QY 26 -----SF 27  
DB 1669 KRFGPAHPECLRTPISDPTMVVWMLHDGKPLEAANRLRMINEFGVCSLDYGVAYSRDGSI 1728  
QY 28 INCKVMSK----- 35  
DB 1729 ITCRATNKYGTGHTSATLIVKDEKSLVVESQLPEGRKGLQRIEELERMAHEGALTGVTTD 1788  
QY 36 -----ALYNRPWR----- 43  
DB 1789 QKEKQKPDIVLYPEPVRVLEGETARFCRVGTGYQPKVNVYINGQLIRKSRFRVRYDGI 1848  
QY 44 -----GLVLSKI----- 50  
DB 1849 HVLDIVDCKSYDTGEVKVTAENPEGVIEHKVLEIQOREDFRSVLRRAPRPEPHVHEP 1908  
QY 51 ----- 50  
DB 1909 GKLOPEQKQVDRPVDTTTETKEVVKLKRAERITHEKVPEESBELRSKFKRTEEGYVEAIT 1968  
QY 51 ----- 50  
DB 1969 AVELKSRKKDSYEBELLRKTKDELHWTKELTEBEKKALABEGKITIPTFKPDKIELSPS 2028  
QY 51 ----- 50  
DB 2029 MEAPKIFERIQSTVGQSDAHFRVVRVVGKDPDCEWYKNGVKIERSDRIYWPEDNVC 2088  
QY 51 ----- 50  
DB 2089 ELVIRDVTAEDSASIMVKAINIAGETSSHAFLVQAKOLITFTQELQDVVAKEKDTMATF 2148  
QY 51 ----- 50  
DB 2149 ECETSEPPFVKVWYKDGMEVHEGDKYRMSDRKHVFLSLITLITDSDAEDYSCVLVEDENV 2208  
QY 51 -----GKYKL 55  
DB 2209 KTTAKLIVEGAVVEFVKELQIEVPESYSGELECIIVPENIEGKWHYHNDVELKSNKXTI 2268  
QY 56 -----DOLK 59  
DB 2269 TSRRGRQNLTVKDVTKEDQGEYSFVIDGKTKTKLKMFRPIAILQGLSDQKVCEDIVQ 2328  
QY 60 LEILRQLET-----TI----- 70  
DB 2329 LEVKSLESVEGVMKDGQEVQPSDRVHIVIDKQSHMLLIEDMTKEDAGNYSFTIPALGL 2388  
QY 71 ----- 70  
DB 2389 STSGRVSVYSDVITPLKDVNVIEGTKAVLECKVSPDVTSVKWLNDQIKPDRDQAI 2448  
QY 71 ----- 70  
DB 2449 VKGTQORLVINRTHASDEGPYKLIIVGRVETNCNLSVEKIKIIRGLDLTCTETQNVFV 2508  
QY 71 -----STKYNVS----- 77  
DB 2509 ELSHSGIDVLWNFKDKIEIKPSSKYKIEAHGKIYKLVLMNMKDDGKYTFYAGENMTSGK 2568  
QY 78 ----- 77  
DB 2569 LTVAGGAISKPLTDQTVAESQBAVFECEVANPDSKGEWLRDGKHLPLTNNIRSESDGHR 2628  
QY 78 -----KQPVKNLTM----- 86



Db 2629 RLIIAATKLDIGEYTYKVATSKTSAKLVKVAVKIKTKLNLFTVTTQDAVFTVELTHEN 2688  
QY 87 -----: : : : :----- 86  
Db 2689 VKGVQWIKNGVWLESNEKVAISVKGTIYSIRIKNCAIVDESUYGFRGLRGASARLHVET 2748  
QY 87 -----: : : : :----- 86  
Db 2749 VKIIKPKDVTALENATVAPEVSVSHDTVPVKWFHKSVEIKPSDKHRLVSRKVKHMLQ 2808  
QY 87 -----NTERPQ----- 92  
Db 2809 NISPSDAGEYTAVGQLECKAKLFPVETLHITKMKNIETPFTKTASFCEVSHFNVPMSW 2868  
QY 93 -----: : : : :----- 92  
Db 2869 LKNGVEIEMSEKPIVVQGLHQLIIMNTSTEDSAEYTFVCGNDQVSATLTVTPIMITSM 2928  
QY 93 -----: : : : :----- 92  
Db 2929 LKIDINAEKDTITFEVTNVYEGISYKWLKNGVEIKSTDKQMRKTKLTHSLNIRNVHFGD 2988  
QY 93 -----: : : : :----- 92  
Db 2989 AADYTFVAGKATSTATLYVEARHIEPRKHIDIKVLEKRAMPECEVSEPDITVQMMKDD 3048  
QY 93 -----YYILAG----- 98  
Db 3049 QELQITDRIKIQEKYVHRLIIPSTRMSDAGKTVVAGGNVSTAKLFVSGRDRVRSIKK 3108  
QY 99 -----: : : : :----- 98  
Db 3109 EVQVIEKORAVFEVNEDDVAHWYKDGIEINFQVQERHKYVWERRIHRMFISETRQSD 3168  
QY 99 -----: : : : :----- 98  
Db 3169 AGEYTFVAGNRSSVTLYNAPPEPOVLOELQVTVQSGKPARFCAMISGRPOPKISWYK 3228  
QY 99 -----: : : : :----- 98  
Db 3229 BEQLSTGPKCKLHDGQEVYLLLIIEAFPEDAAYTCEAKNDYGVATTSAVVEVPEV 3288  
QY 99 -----PION----- 102  
Db 3289 SPDQEMPVYPPAIITPLQDVTVSEGQPARFCQVSGTDLKVSWSKDKKIKPSRPPRMTQ 3348  
QY 103 -----: : : : :----- 102  
Db 3349 FEDTYOLEIAEAYPEDEGTYTFVANNVQVSGSTANLSLEAPESILHERIEQEIEMEMKE 3408  
QY 103 -----: : : : :----- 102  
Db 3409 FSSSFLSABEGLHSAELQSKINETLELLSSBPVYTPKDESEKGTGPIFIKEVSNADI 3468  
QY 103 -----: : : : :----- 102  
Db 3469 SMGDVATLSVTVIGIPKPIQWFFNGVLLTPSADYKVFVFGDDHSLIILFTKLEDBEY 3528  
QY 103 -----: : : : :----- 102  
Db 3529 CMASNDYGKTI CSAYLKINSKGEHGDTESETSAVAKSLEKLGCGCPPHFLKELKPIRCAQ 3588  
QY 103 -----YSL-----TYLWF----- 110  
Db 3589 GUPAIFEYTVGPEPATVTFWFKENKQCTSVYVYTIHNPNGSGTFIINDPQDESDGLYIC 3648  
QY 111 -----DFYSTQLRKPA----- 121  
Db 3649 KAENMLGESTCAEALLVLEDTDMTDPCKAKSTPEAPDFQTPKLGPAVEALDSEQEI 3708  
QY 122 -----: : : : :----- 121

Db 3709 ATPVKDITLKAALITEENQOLSIEHIAKANELSSQLPLGAQELQSLILEODKLTPESTRFP 3768  
QY 122 -----: : : : :----- 121  
Db 3769 LCINGSIHFOPLKEPSNQLQIVOSQKTFKSEGILMPPEPETAVALSDTEKIFPSAMSI 3828  
QY 122 -----: : : : :----- 121  
Db 3829 EQINSLTVEPLKTLAABPEGNYQSSIBPPMHSYLTSAEVLSLKERTVSDTNRQRVT 3888  
QY 122 -----: : : : :----- 121  
Db 3889 LQKQEAQSALIILSQAEGHVESLOSVDVMSIQVNYEPLVPSEHSCTEGKILIESANPL 3948  
QY 122 -----: : : : :----- 121  
Db 3949 ENAGQDSAVRIEBEGKSLRPLALEBKQVLLKEHSDNVVMPDQIIESKREPVAIKKVOE 4008  
QY 122 -----: : : : :----- 121  
Db 4009 VQGRDLLSKESLLSGIPEBQRLNLKIQICRALQAAVASQPLFSEWLRNIEKVEAVN 4068  
QY 122 -----: : : : :----- 121  
Db 4069 ITQEPRHIMCYLTVTSKASVTEVTIIIEDVDPMANLKMELDALCALIYBEIDILTAE 4128  
QY 122 -----: : : : :----- 121  
Db 4129 GPRIQQAKTSLQEBMDSPSGQKVEPIPEPEVSKYLSTEBVSYFNQSRVKYLDATP 4188  
QY 122 -----: : : : :----- 121  
Db 4189 VTKGVASVVSDEKQESLKPSEKSESSESGTBEVATVKIQEABGGLIKEDGPMIHTP 4248  
QY 122 -----: : : : :----- 121  
Db 4249 LVDTVSEGDIVHLTTSITNAKEVNWYPENKLVPSDEKFKCLQDQNTYTLVIDKVTEDH 4308  
QY 122 --KYVYSQYNHTAKTI----- 135  
Db 4309 QGEYVCEALNDSGKTSATSAKLVVKKRAAPVIRKRIEPLVALGHLAKFTCEIQSAPNVR 4368  
QY 136 -----: : : : :----- 135  
Db 4369 QWFKAGREIYESDKSIRSSKYISSLEILRTQVDCGYTCKASNEYSVGSCTATLTVT 4428  
QY 136 -----: : : : :----- 135  
Db 4429 PGGEKKVRKLLPERKPEKPEEVVLSVLRKRPEEBEPKVEPKKLEKVKPAVPEPPPKP 4488  
QY 136 -----TFRPPP----- 141  
Db 4489 VBEVEVPTVKRERKIPEPTKYPEIKPAIPALPAPKPKPEAEVKTIKPPVPEPEPTPIA 4548  
QY 142 -----: : : : :----- 141  
Db 4549 APVTVPVVGKAAKAPKEAAKPGPIKGVPKTPSPIEABRRKLRPGSGGKPPDEAP 4608  
QY 142 -----: : : : :----- 141  
Db 4609 FTYQLKAVPLKFKVKEIKDIIILTESEBFVSSAIFECVLSPSTAITTMMKOGSNIRESFKIR 4668  
QY 142 -----: : : : :----- 141  
Db 4669 FIADGDKRKLHIIDVQLSDAGEYTCVLRGNKEKTSTAKLVVEELPVRPVKLTLEBEVTV 4728  
QY 142 -----C-----GRV----- 145  
Db 4729 KGQPLYLSCELNKERDVVVRKDGKIWEKPGRIVPFGVIGLRAALTINDADTDAGTYTVT 4788  
QY 146 -----: : : : :----- 145  
Db 4789 VENANNLECCSVCVKVVEVIRDWLVKPIRQHVKPGTAIFACDIAKDTNIRKFKGYDEI 4848

QY 146 ----- 145  
Db 4849 PAEPNDKTEILRDGNHLYLKIKNAMPEDIAEVAVEIEGKRYPAKLTIGEREVELLKPIED 4908  
QY 146 -----PSMTC----- 150  
Db 4909 VTIYEKESAFDAIESEADI PGOWKLKGELLRPSPTCEIKABGGKRFLLTHVKVLDQAGE 4968  
QY 151 ----- 150  
Db 4969 VLYQALNAITTAITLVKEIELDAVPLKDVTPERRQARFECVLTREANVINSGPDIK 5028  
QY 151 ----- 150  
Db 5029 SSDKFDIIADGKHILVINDSQDDDEGVYTAIEVEGKTSARLFTVGTIRLKFMSPLEDQTV 5088  
QY 151 ----- 150  
Db 5089 KEGETATFVCELSHERKHVVWFKNDAKLHTSRVTLISSEBKTTHLEMKVETLDDISQIKA 5148  
QY 151 ----- 150  
Db 5149 QVKELSSAQKLVLEADPYTVTKLHDKTAVEKDEITLKCEVSKDVPVKWFKDGEIIVPSP 5208  
QY 151 ----- 150  
Db 5209 KYSIKADGLRILKI KKA DLKDXGEVYCDGTDKTKANVTVEARLIEVEKPLYGVEVFG 5268  
QY 151 ----- 150  
Db 5269 ETAHFEILSEPDVHGOWKLKGQPLTASPDCEIIEDGKKHILHNCQLGMTGEVSFQAA 5328  
QY 151 ----- 150  
Db 5329 NAKSAANLKVKEPLIFITPLSDVKVFEKDEAKFECEVSREPKTFRWLKGTOBITGDDRF 5388  
QY 151 ----- 150  
Db 5389 ELIKDGTKHSWVIXSAFDEAKYFEADKHTSGKLIIEGIRLKFTPLKDVTAKEKES 5448  
QY 151 ----- 150  
Db 5449 AVFTVELSHDNIRVKWFKNDQRLHTRSVSMODEGKTHSITFKDLSIDDTSQIRVEAMGM 5508  
QY 151 ----- 150  
Db 5509 SSEAKLTVLESDPYFTGKLQDYGVEKDEVILQCEISKADAPVKWFKDGEIKEPSKNAVI 5568  
QY 151 ----- 150  
Db 5569 KTDGKKEMLILKKA LKSDIGQYTCDCGTDKTSGLKDIEDREIKLVRPLHSVEVMEETAR 5628  
QY 151 ----- 150  
Db 5629 FETEISEDDIHANWKLKGEALLOTPDCEIKEEGKIHSLVHLNCRLDQGTGVDFQAANVKS 5688  
QY 151 ----- 150  
Db 5689 SAHLRVKPRVIGLLRPLKDVTTAGETATPDCELSYEDI PVWYVLKGLKLEPSDKVVPRS 5748  
QY 151 ----- 150  
Db 5749 EGKVHTILTRDVKLEDAQEVLTAQDKFTHANLPVKEPPVEFTKPLEDQTVBEGATAVLE 5808  
QY 151 ----- 150  
Db 5809 CEVSRENAKWKFKNGTEILSKKYEIVADGRVKLVHDCPTBEDIKTYTCDAKDKFTSC 5868  
QY 151 ----- 150  
Db 5869 NLNVVPHVEFLRPLTLQVREKEMARFECELSRENAKWKWFKDGAELKKGKYYDIISKG 5928

QY 151 ----- 150  
Db 5929 AVRILVINKCLLDDEAEYSCVETARTSGMLTVLEBAVFTKNLANIEVSETDTIKLVCE 5988  
QY 151 ----- 150  
Db 5989 VSKGAEVWYKGBDEEIIETGRYBILTEGRKRIILVIONAHLEDAGNYNCRPLSPSRTDGKV 6048  
QY 151 ----- 150  
Db 6049 KVHELAAEFISKPNLEILEGEKAEFVCSISKESFPQWKRRDDKTLES GDKYDVIA DGK 6108  
QY 151 ----- 150  
Db 6109 RVLVVKDATLQDMGTYYVMVGARAAAHLTVIEKLRIVVPLKDTRVKESQEVWFNCEVNT 6168  
QY 151 ----- 150  
Db 6169 EGAKAKWFRNEEAFDSSKYIILQKDLVYTIIRDALHDDQANYVSLTNHGENVKSA 6228  
QY 151 ----- 150  
Db 6229 NLIVEEEDLRIVEPLKDIETMEKKSVTFCWKVNLNVTWKWTKNGEEVFPDNRSYRVDK 6288  
QY 151 ----- 150  
Db 6289 YKHLMTIKDCGFPDRGEYIVTAGODKSVABLLIIEAPTEFVEHLEDQTVTFEDDAVFCQ 6348  
QY 151 ----- 150  
Db 6349 LSREKANVWYRNGREIKEGKYKFEKDGSHRLI IKDRLDDBCEYACGVEDRKSARL 6408  
QY 151 -----LSEM----- 154  
Db 6409 FVEEIPVEIIRPPODILEAPGADVVFVLAELNKDKVEQVQLRNNMNVVQGDQKHMSEGI 6468  
QY 155 ----- 154  
Db 6469 HRLQICDKPRDQGEYRFIAKDKEARAKLELAAPKIKTADODLVVDVVKPLTMVVPYDA 6528  
QY 155 ----- 154  
Db 6529 YPKAAEFKENEPLSTKTIDTTABQTSFRILEAKKGDKGRYKIVLQNKHGKAEGLNLK 6588  
QY 155 ----- 154  
Db 6589 VIDVPVRNLEVTETFDGEVSLAWEEPLTDGSKIIGYVVERRDIKRKTWYLATDRAES 6648  
QY 155 -----LNV----- 157  
Db 6649 CEFTVTGLQGGVEYLFVRSARNRVGTGEVETDNPVEARSKYDVPGPLNTIITDVNRF 6708  
QY 158 ----- 157  
Db 6709 GVSLTWPEPYDGAETINYVIELRDKTSIRWDTAMTVRAEDLSATVTDVBEQYESFRV 6768  
QY 158 ----- 157  
Db 6769 RAQNRIGVGPSAATPFVKVADPIERPSPVNLTSDDQTSVQLKWEPLKDGSGPILG 6828  
QY 158 ----- 157  
Db 6829 YIIERCEGKNWIRCNMKLVPELTYKVTGLEKGNKYLVRSAENKAGVSDPSEILGPLT 6888  
QY 158 ----- 157  
Db 6889 ADDAFVEPTMDLSAFKDGLEIVIPNPIITILVPSTGYPRPTATWCFGDKVLETGDRVKMT 6948  
QY 158 ----- 157  
Db 6949 LSAYABLVISPSERSDKGIYTLKLENRVKTIISGEIDVNVVIARPSAPKELKFGDITKDSVH 7008  
QY 158 ----- 157

Db	7009	LWEPDDDDGGSLTGYYVKEKREVSRTWTKVMDFVTDLEFVTPDLVQKKEYLKVCA RN	7068	Db	8089	VEVHNPTAEAMTITWKPPPLYDGGSKIMGYIIIEKIAKEBERWKRCEHNLVPILTYYTAKGLE	8148
Qy	158	-----	157	Qy	170	-----	169
Db	7069	KCGGEPAYVDEPNMSTPATVPDPENVKWRDRTANSIFLTWDPKNDGGSRIGKIYIVE	7128	Db	8149	EGKEYQFRVRAENAAGISEPSRATPPTKAVDPIDAPKVILRTSLEVKRGDEIADASISG	8208
Qy	158	-----	157	Qy	170	-----	169
Db	7129	RCPRGDKWVACGEPVAETKMEVTLGLEEGKYAYRVKTLNRQGASKSPRTBIEQAVDTQ	7188	Db	8209	SPYPTITWIKDENVIVPEEIKKRAAPLVRRRKGEVQEBBPFVLPTQRLSIDNSKKGESQ	8268
Qy	158	-----	157	Qy	170	-----	169
Db	7189	EAPFELDVKLKLAGLVKAGTKIELPATVTGKPEPKITWTKADMILKQDKRITNIENVPKK	7248	Db	8269	LVRDSLRLPDHGLYMIKVENDHGIAPACTVSVLDTPGPPINPFVFDIRKTSVLCKWEPP	8328
Qy	158	-----	164	Qy	170	-----	172
Db	7249	STVTIVDSKESDGTGTYIEAVNVGRATAVVENVLDKCPAPAFDITDVTNESCILLTNW	7308	Db	8329	LDDGGSEIINYLTLEKDKTKPDSEMIWVTSTLRHKYSVTKLIEGKEYLFRVRAENRFGP	8388
Qy	165	-----	164	Qy	173	-----	178
Db	7309	PPRDDGSKIYVVERRATDSBWMHKLSTVKDTNFKATKLI PNKEYIFRVAENNYGA	7368	Db	8389	GPPCVSKPLVAKDPFGPPDAPDKPIVEDVTSNMLVKWNEPKDNGSPILGYWLEKREVNS	8448
Qy	165	-----	164	Qy	179	-----	178
Db	7369	GEVQASPIAKYQFDPGPGPTLRLPSDITKDAVTLTWCEPDDGGSPITGYWVERLDPD	7428	Db	8449	THWSRVNKSLLNALKANVDGLLEGLTYVFRVCAENNAAGPKFSPSPDPKTAHDPISPSPGP	8508
Qy	165	-----	164	Qy	179	-----	178
Db	7429	TDKWVRCKMPVKDITYRVKGLTNKKYRFRVLAENLAGPKSKSTEPILIKDIPDPW	7488	Db	8509	PIPRVTDTSSTIELEWEPPAFNGGGEIVGYFVDKQLVGTNKNWSRCTEKMIKVQYTVKE	8568
Qy	165	-----	164	Qy	179	-----	178
Db	7489	PPCKPTVKDVGKTSVRLNWKTPHEDGGAKIESVYIEMLKTGTDWVRVAGVPTTQHLLP	7548	Db	8569	IREGADYKLRVSAVNAAGEGPGGETQPVTVABPQEPPEPPAVELDVSKGGI QIMAGKTLRIP	8628
Qy	165	-----	164	Qy	179	-----	178
Db	7549	GLMEGOYSFRVRAVNKAGESESPSPVLCREKLYPPSPRWLEVINITKNADLKWT	7608	Db	8629	AVVTGRPVTKWTKBEGELDKORVIDNVGKSELIIKDALRKDHGRVYVITATNSCGSK	8688
Qy	165	-----	164	Qy	179	-----	178
Db	7609	VPEKGGSPITNIVKRDVRRKGWQTVDTTVKDTKCTVPLTEGSLYVFRVAENAIQ	7668	Db	8689	FAARVEFDPVGPVLDLKPVVTRNKMCLLNWSDEDDGGSEITGPIIERKOKAMHTWRQ	8748
Qy	165	-----	164	Qy	179	-----	178
Db	7669	SDYTEIEDSVLAKDTFTTGGPVVALAVDVTKRHDVKWEPKNDGGRPIQRVIEKKER	7728	Db	8749	PIETERSKCDITCLLEGQYKFRVIAKNKFGCGPPVEIGPILAVDPLGPTSPERLTYTE	8808
Qy	165	-----	169	Qy	179	-----	178
Db	7729	LGTRWYKAGTAGDCNFRVTDVIEGTEVQFQVRAENEAGVGHPSPTBILSIEDTSP	7788	Db	8809	RORSTITLDWKEPRNGSGSPIQGYIIIEKRRHDKPDFERVNKRCLCPTTSPLENLDEHMY	8868
Qy	170	-----	169	Qy	179	-----	178
Db	7789	SPPLDLHTVTDAGRKHIAIAWKPEKNGGSPITGYHVEMCPVGTCKMVRVNSRPIKDLKPK	7848	Db	8869	EFRVKA VNEIGSEPSLP LNVIQDDEVPTTIKLRLSVRGDTIKVKAGEPVHVPADVTGL	8928
Qy	170	-----	169	Qy	179	-----	178
Db	7849	VBEGVVPDKEYVLRVRAVNAIGVSEPSSEISENVVAKDPCKPTIDLETHDII VIEGEKLS	7908	Db	8929	PMPKIEMSKNETVIEKPTDALQITKEEVSRSBAKTELSIPKAVREDKGTGTYTVTASNRLGS	8988
Qy	170	-----	169	Qy	179	-----	178
Db	7909	IPVFRVAVPVTVSMHKGKVKASDRLTMKNDHISAHLEVPKSVRADAGIYITILENKL	7968	Db	8989	VFRNVHVEYDRPSPPRNLA VTDIKAESCYLTDAPLDNGGSEITHYVIDKRDASRKKAE	9048
Qy	170	-----	169	Qy	179	-----	178
Db	7969	GSATASINVKVLGPGCKDIKASDITKSSCKLTWEPPEFDGTPILHYVLERREAGR	8028	Db	9049	WEVNTTAVEKRYGIWKLI PNQOYEFVRVAVNKYGISDECKSDKVVIQDPYRLPGPPGPK	9108
Qy	170	-----	169	Qy	179	-----	181
Db	8029	YIPVMSGENKLSWTVKDLIPNGEYFFRVRVAVNKVGGGEIELKNPVIAQDPKQPPDPVD	8088	Db	9109	KVLARTKGSMLVSWTPTPLDNGGSPITGYWLEKREBGS PYWSRVSRAPITKVGLKGVEFNV	9168
Qy	170	-----	169	Qy	182	PR-----	183
Db				Db	9169	PRLEGVKYQFRAMAINAAGIGPPSEPDPVAGDPFIPPPPPSPCEVKDKTKSSISLGW	9228

QY 184 ----- 183  
Db 9229 KPPAKDGGSPKIGYIVEMOEQTDMKRVNEPDKLITTCBCEVVPNLKELRYFRVKAVN 9289  
QY 184 ----- 183  
Db 9289 EAGESPSTTGEIPATDIOEBEPVIDIGAODCLVCKAGSQIRIPAVIKGRPTPKSSWE 9348  
QY 184 ----- 183  
Db 9349 FDGAKKAMKGVHDIPEDAQLEAENSSVIIPECKRSHTKGYSITAKNKAGQKTANCR 9408  
QY 184 ----- 183  
Db 9409 VKVMDVPGPPKOLKVSDITRGSCRLSWKMPDDGGDRIKGYVIEKRTIDGKAWTKVNPDC 9468  
QY 184 ----- 183  
Db 9469 GSTTFVVDLLSEQYFFRVAENRFGIGPPVETIORTTARDPIYPPDPPIKIKIGLITK 9528  
QY 184 ----- 183  
Db 9529 NTVHLSWKPPKNDGSPVTHYIIVECLAWDPDTGTKEARQCNRDVEELOFTVEDLVEGG 9588  
QY 184 ----- 183  
Db 9589 EYEFVRKANNAAGVSKPSATVGPCDCQRPDMPPSIDLKEFMEVEEGTNVNIIVAKIKGVFP 9648  
QY 184 ----- 183  
Db 9649 PTLTWFKAPKPKDNKEPVLVDTHVNKLVDVDDCTLVI PQSRSDTGLYITITAVNNLGTA 9708  
QY 184 ----- 183  
Db 9709 SKEMRLNVLRGPPVGPPIKFESVSADQMTLSWPPKDDGSKIITNVIKREANRKTWV 9768  
QY 184 ----- 183  
Db 9769 HVSSEPKECTYTI PKLLEGHEYFRIMAQNKYIGIGEPDSEPETARNLFSVPGAPDKPTV 9828  
QY 184 ----- 183  
Db 9829 SSVTRNSMTWBEPEYDGGSPVTGYWLEMDTTSKRKRVNRDPKAMTLGVSYKVTGL 9888  
QY 184 ----- 183  
Db 9889 IEGSDYQFRVYAINAAGVPASLPSPDATARDPIAPGPPFPKVTWTKSSADLEWSPPL 9948  
QY 184 ----- 183  
Db 9949 KDGSKVTGYIIVEYKEGKEWEKGDKEVRGTKLVVTGLKEGAFYKFRVSAVNIAGIGE 10008  
QY 184 ----- 183  
Db 10009 PGEVTDVIEMKDLRVSPDLQDASVRDRI VVHAGGVIIRIIAYVSGKPPPTVTWNMERTL 10068  
QY 184 ----- 183  
Db 10069 PQEATIETTAISSMWINKCORSHQGVYSLAKNEAGERKTIIVDVLDPGVPGPFLA 10128  
QY 184 ----- 183  
Db 10129 HNLTNESCKLTFWSPDEDDGGSPITNVIKESDRRAWTPVTVTTRQNAVQGLIQGKA 10188  
QY 184 ----- 191  
Db 10189 YFFRIAAENSIGMGFFVETSEALVIREPITVPERPEDLEVKEVTNKTVTLTWNPCKYDGG 10248  
QY 192 ----- 191  
Db 10249 SEIINYVLESRLIGTEKHFKVTNDNLLSRKYTVKGLKEGDTEYVRVSAVNIIVGQKPSFC 10308

QY 192 ----- 191  
Db 10309 TKPITCKDELAPPTLHLDFRDKLTIRVGEAFALTGRYSGPKPKVSWFKDEADVLEDDRT 10368  
QY 192 ----- 191  
Db 10369 HIKTTPATLALAKIKAKRSDSGKYCVVVENSTGSRKGFQVNVVDHPGPPVGPVSFDEVT 10428  
QY 192 ----- 191  
Db 10429 KDMVVISWKBPPLDDGSGKITNYIIIEKKEVGKVMMPVTSASAKTTCKVSKLLEGKDIYFR 10488  
QY 192 ----- 191  
Db 10489 IHAENLYGISDPLVSDLSMKAKDRFRVPDAPDQPIVTEVTKDSALVTWNKPHDGGKPTNY 10548  
QY 192 ----- 191  
Db 10549 ILEKRETMKSKWARVTKDPIHPYTKFRVPDILLEGQCYEFRVSAENEIGIDGPPSPKPVF 10608  
QY 192 ----- 191  
Db 10609 AKDPIAKPSPVPNPEADITTCNSVDLTWQPPRHDGSKILGYIIVEYQKVGDDEWRANHT 10668  
QY 192 ----- 191  
Db 10669 PESCPETKYKVTGLRDGQTYKFRVLAVNAAGESDPAHVPEVLVKDLRLEPPELILDANMA 10728  
QY 192 ----- 198  
Db 10729 REQHIKVGDTLRLSAI IKGVFPFKVTKKEDAPTKARIDVTPVGSKLEIRNAAHEDGG 10788  
QY 199 ----- 198  
Db 10789 IYSLTVENPAGSKTVSVKVLVDKPGPRDLEVSIRKDSCLTWLKEPLDDGGSVITNYV 10848  
QY 199 ----- 198  
Db 10849 VERRDVASAQMSPLSATSKKSHFAKHLNEGQYLFVAAENQYGRGPFVETPKFIKALD 10908  
QY 199 ----- 198  
Db 10909 PLHPPGPKDLHHVDVDKTEVSLVWNKPRDRDGGSPITGVLVEYQEGTQDWIKFKVTNVL 10968  
QY 199 ----- 207  
Db 10969 ECVWTGLOQGTKYFRFRVKAENIVGLGLPDTTIPIBCQEKLVPPSVELDVKLEGLVVRAG 11028  
QY 208 ----- 207  
Db 11029 TTVRFPALIRGVPVPTAKWTTDGSBIKTDHYTVETDNFSSVLTIKNCLRRDTGEVQITV 11088  
QY 208 ----- 207  
Db 11089 SNAAGSKTVAVHLTVLDPGPTGPINILDVTPHEMTISWQPPKDDGGSPVINIYVEKOD 11148  
QY 208 ----- 219  
Db 11149 TRKDTMGVVSSGSGSKTKLIPHLOKGEYVFRVRAENKIGVGPPLDSTPTVAKHKFSPPS 11208  
QY 220 ----- 219  
Db 11209 PFGKPVVTDITENAAVSWTLPKSDGSGSPITGYYMERREVTGKWRVANKPTIADLKFRVT 11268  
QY 220 ----- 219  
Db 11269 GLYEGNTYFRVFAENAGLSKSPSSDPIKACRPKPPPINPKLKDOKSRETADLVWT 11328  
QY 220 ----- 219  
Db 11329 KPLSDGSGPILGVVVEQKPGTAQWNRINKDELIRQCAPRVPGLEEGNEYFRIKAANIV 11388  
QY 220 ----- 219

Db 11389 GECEPRELAESVJAKDILHPPPEVELDVTCRDVITVRVGOTIRILARVKGRPEPDITWTKE 11448  
QY 220 ----- 219  
Db 11449 GKVLVREKRVLDLIQDLPRVELQIKEAVRADHGKYIISAKNSSGHAQSAIVNVLDPRGPC 11508  
QY 220 ----- 219  
Db 11509 QNLKVTNVTKENCTISWENPLDNGSSEITNFIVEYRKNQKGSIVASDVTKRLIKANLL 11568  
QY 220 ----- 219  
Db 11569 ANNEYFRCAENKVGVGPTIETKTPILAINPIDRPGEPENLHDKGTFVYLKWRPD 11628  
QY 220 ----- 219  
Db 11629 YDGGSPNLSYHVERRLKGDDMERVHKGSIKETHYVMDRCVENQIYEFPRVQTKNEGSED 11688  
QY 220 ----- 219  
Db 11689 WVKTEVVVKEDLQKPVLDKLSGVLTVKAGDITRLEAGVRGKPFPEVAWTKDKOATDILT 11748  
QY 220 ----- 219  
Db 11749 RSPRVKIDTRADSSKPSLTKAKRSDGGKYVVVATNTAGSFVAVATVNVLDKPGFVRNLKI 11808  
QY 220 ----- 219  
Db 11809 VDSSDRCTVCWDPPEDDGCETQNYILEKCEKRMVWSTYSATVLTGCTTVTRLIEGNE 11868  
QY 220 ----- 219  
Db 11869 YIFRVAENKIGTPTESKPIAKTKYDKGRDPPEVTKVSKEMTVMNPPEDYDGGK 11928  
QY 220 ----- 219  
Db 11929 SITGYLEKKEKSTRVWPVNKSAIPERRMKVONLLPDHEYQPRVKAENEIGIBEPSLPS 11988  
QY 220 ----- 219  
Db 11989 RPVWAKDPIEPGPPNFRVVDTKHSITLHGKPVYDGGAPFIIGVVEWRPKIADASPD 12048  
QY 220 ----- 219  
Db 12049 EGWKRCAAAQLVRKEFTVTSLDENQYEFVRVCAQNVGIGRPAELKEAIKPKBILEPPE 12108  
QY 220 ----- 219  
Db 12109 IDLDASMRKLVIVRAGCPIRLFAIVRGRPAKVTWRKVGIDNVVRKGQVDLVDTWAPLVI 12168  
QY 220 ----- 219  
Db 12169 PNSTRDSSKYSLTLVNPAGEKAVFVNVRLDTPGVPVSLKVSDDVTKTSCHVSWAPPEND 12228  
QY 220 ----- 227  
Db 12229 GGSQVTHYIVEKREADRKTWSTVTPVKKTSFHVTLNVPGENYFRVAVNEVPGVPTD 12288  
QY 228 ----- 227  
Db 12289 VPKPVLASDPLSEDPDPKLEATEMTKNSATLAWPLPLRDGGAKIDGYIISYREEBPAD 12348  
QY 228 ----- 227  
Db 12349 RWTEYSVVKDLSLVVTGLKEGKYKFRVAARNVAVGSLPREAGVYEAKEQLLPPKILMP 12408  
QY 228 ----- 227  
Db 12409 EQITIKAGKRLIEAHVYKPHPTCKWKKGEDSVTSSHLAVHKADSSSILIIKDVTRKD 12468  
QY 228 ----- 227

Db 12469 SGYISLTAENSSGTDTQKIKVVVMDAPGPPQPPFDISDIDADACSLSWHIPLEGGSNIT 12528  
QY 228 ----- 227  
Db 12529 NVIVEKDVSRGDMWTALASVTKTSRUGKLIPIQGEYIIFRVAENRFGISEPLTSPKWA 12588  
QY 228 ----- 227  
Db 12589 QPFFGVPSEPKNARVTVKNKDCIFVWDRPDSGGSPIIIGYLIERKERNLLWVKANDTL 12648  
QY 228 ----- 227  
Db 12649 VRSTEYPCAGLVEGLEYSFRIYALNKAAGSSPPSKTEYVTARMPVDPGKPEVIDTKST 12708  
QY 228 ----- 227  
Db 12709 VSLIWARPKHDGSKIIGYFVEACKLPGDKWVRCNTAPHQIQEEYATATGLEBAQYQFR 12768  
QY 228 ----- 227  
Db 12769 AIARTAVNISPESEPDVPTILAENVPRIDLSVAMKSLLTVKAGTNVCLDATVFGKMP 12828  
QY 228 ----- 233  
Db 12829 TVSWKDKDGTLLKPAEGIKWAMORNLCLELFSVNRKDSGDIYITAEENSSGKSATIKLV 12888  
QY 234 ----- 233  
Db 12889 LDKPGPASVINKMWSDRAMLSWBPLEDGGSEITNYIVDKRETSRPNWAQVSATVPIT 12948  
QY 234 ----- 233  
Db 12949 SCSEKLIIEGHEYQFRIKAENKYGVDPVFTBPAIAKNPYDPGRCDDPPVISNITKDHT 13008  
QY 234 ----- 233  
Db 13009 VSWKPPADGGSPITGYLLEKRETOAVNWKVNRKPIERTLKATGLQEGTEYEFRTAI 13068  
QY 234 ----- 233  
Db 13069 NKAGPKPSDASKAAVARDPQYPPAPPAPKVVYDTRSSVSLSWGKPAYDGGSPIIIGYL 13128  
QY 234 ----- 233  
Db 13129 EVKRADSNWVRNCLPQNLQKTRPEVTGLMEDTQYQFRVYAVNKIGYSDPSDVPDKHYK 13188  
QY 234 ----- 233  
Db 13189 DILIPPEGEHDADLRKTLILRAGVTMRLYVPVKGRPPPKITWSKPNVNLDRIGLDIKST 13248  
QY 234 ----- 233  
Db 13249 DFDTLRCENVNKYDAGKYILLTLENSCGKKEYTIVVKVLDTPGPPINVTVKEISKDSAV 13308  
QY 234 ----- 233  
Db 13309 TWEPIIDGGSPIINYVQKRAERKSWSTVTTEGKTSFRVPLEEGSKSYFRVFAENE 13368  
QY 234 ----- 233  
Db 13369 YGIGDPGETRDAVKASQTPGPVVDLKVRSVSKSGSIGWKPHSGGSRIIIGYVVDFLTE 13428  
QY 234 ----- 233  
Db 13429 ENKQVRKMSLSLOYSAKDLTEGKEYTFRVSAENENGEBTPSEITTVARDDVVAPDLDL 13488  
QY 234 ----- 233  
Db 13489 GLPDLCLAKENSFRLKIPKNGKAPSVSWKKGBDPLATDTRSVSESSAVNTTLIVYDC 13548  
QY 234 ----- 233  
Db 13549 QKSDAGKYTITLKNVAGTKEGTISIKVVGKPGIPTGPIKDFEVTAEAMTLKWA PKDDGG 13608

QY	234	-----	233	QY	262	-----	267
Db	13609	SBITNYLEKRDSDVNNKWTCASAVQKTTFRVTRLHEGMEYTFRVAENKYGVEGLKSE	13668	Db	14689	DGGVPISNVYVEMRQDSTTWELATTVIRTYYKATRLTTLGLEQFRVKAQNRYGVGPGI	14748
QY	234	-----	233	QY	268	-----	267
Db	13669	PIVARHPFDVDPAPPNNIVDRHDSVSLTWDPKKTGGSPITGYHLEPKERNLLWKRA	13728	Db	14749	TSAWIVANYPFKVPGPPTQVTAVTKDGMTISWHBPLSDGGSPILGYHVERKERNILW	14808
QY	234	-----	236	QY	268	-----	267
Db	13729	NKTPIRMDFKVTGLTEGLEFRVMAINLAGVGKPSLPBPVALDPIDPFGKEVINI	13788	Db	14809	QTVSKALVPGNIPKSSGLTDCGIAEFVRIAENMAGKSPKSEPMALALDIPDPGKVP	14868
QY	237	-----	245	QY	268	-----	267
Db	13789	TRNSVTLIWTEPKYDGGHKLGTGYIVEKRDLPKSMKANVNVPECAFTVTDLVEGGKYE	13848	Db	14869	LNITRHTVTLKWAPEYTGFGFKITSYIVEKRDLPNGRWLKNFNSNILENEFTVSGLTEDA	14928
QY	246	-----	245	QY	268	-----	267
Db	13849	FRIRAKNTAGASAPSESTETIICKDEYEAPTIVLDPITIKDGLTIKAGDTIVLNAISILG	13908	Db	14929	AYEFRVIAKNAAGAISSPSEPSDAITCRDDVEAPKIKVDVKFKDTVILKAGEAFRLADV	14988
QY	246	-----	245	QY	268	-----	267
Db	13909	KPLPKSSWAKAGDIRPSDITQITSTPTSSMLTIKYATRKDAGEYITATNPFGTKVHV	13968	Db	14989	SGRPPPTMEWSDGKLEGTAKLEIKIADFSTNLVNKDSRTRDSGAVTLTATNPGGFAKH	15048
QY	246	-----	245	QY	268	-----	267
Db	13969	KVTLDVPGPGPVEISNVSAEKATLTWTPPLEDGGSPIKSYILEKRETSRLLWTVVSSED	14028	Db	15049	IFNVKVLDRPGPBGPLAVTEVTSKCVLSWFPPLDDGGAKIDHYIVQKRETSRLATNV	15108
QY	246	-----	K 246	QY	268	-----	267
Db	14029	IQSCRHVATKLIQNEVI FRVSAVNHVGKEPVQSEPVKMDRFGPPGPEKPEVSNVTK	14088	Db	15109	ASEVOVTKLVTKLLKMGNEYIFRVMAVNVKYGVEPSEPVIAVNPYPPDPKPNVTT	15168
QY	247	NT-----	253	QY	268	-----	270
Db	14089	NTATVSKRPVDDGGSEITGYHVERBKSLRWRAIKTPVSLRCKVTGLQEGSYEPR	14148	Db	15169	ITKDSMVVCGHPDSDGSEIINVIYERRDKAGORWIKCNKKTTLTDLRYKVSGLTEGHEY	15228
QY	254	-----	253	QY	271	-----	278
Db	14149	VSAENRAGIOPPSEASDSVLMKDAAYPPGPPSNPHVTDTTKSASLAWGKPHYDGLLEIT	14208	Db	15229	EFRIMAENAGISAPSPSPFYKACDVTVPKPPGNPRVLTSSRSISIAWNKPIYDGS	15288
QY	254	-----	253	QY	279	-----	278
Db	14209	GYVVEHQVGEAWIKOTTGTALRITQFVVVDLQTKENYFRISAINDAGVGEPAVDPV	14268	Db	15289	EITGMVEIALPEDEWQIVTPPAGLKATSYITIGLTENQEKIRIYAMNSEGLGEPALV	15348
QY	254	-----	258	QY	279	-----	278
Db	14269	EIVEREMAPDFELDAELRRLVVRAGLSIRIFVPIKGRPAPEVTTWKDNLNKRANEN	14328	Db	15349	PGTPKAEDRMPLPPIELDADLRKVVTIRACCTLRLFVPIKGRPDPEVKWARDHGESLDA	15408
QY	259	-----	258	QY	279	-----	281
Db	14329	TESFTLLIIPECNRYDTGKFVMTIENPAGKSGFVNVRLDTPGPVNLNRPDTDKDSVT	14388	Db	15409	SIESASSYTLIIIVGNVNRFDGKYILTVENSNGSKSAFNVVRVLDTPGPPQDLKVKEYTK	15468
QY	259	-----	258	QY	282	-----	281
Db	14389	LHWDPLIDGSRITNYIVEKREATRKSYSTATTCKHCTYKVTGLSEGCEYFRVMAEN	14448	Db	15469	TSVTLTWDPPLDGGSKIKNYIVEKRESTRKAYSTVATNCHTKTSMKVQDQBGCSYFVRV	15528
QY	259	-----	258	QY	282	-----	281
Db	14449	EYIGEPTEPTEPVKASEAPSPDLSLIMDITKSTVSLAWPKPKHDGSKITGYVIEAQR	14508	Db	15529	LAENEYIGLPAETAESVKASERPLPGKITLMDVTRNSVLSWBKPEHDGGRILGYIV	15588
QY	259	-----	261	QY	282	-----	289
Db	14509	KGSDQWTHITTVKGLCVVRNLTEGEBYTFQVMAVNSAGRSAPRESRPVIVKQTMLEL	14568	Db	15589	EMQTKGDKWATCATVVKVTEATITGLIQGEEYSFRVSAQNEKGISDPRQLSVPIAKDLV	15648
QY	262	-----	261	QY	290	-----	289
Db	14569	DLRGYOKLVIKAGDNKIVEIPVLRPKPTVTTWKGDQILKQTORVNFETTATSTILNI	14628	Db	15649	IPPAFKLLFNTFTVLAGEDLKVDPFPIGRPTPAVTWHKONVPLKQTTVRNAESTENNSLL	15708
QY	262	-----	261	QY	290	-----	289
Db	14629	NECVRSDSGPPLTARNIVGEVGDVITIQVHDIIPGPTGPIKPDDEVSSDFVTTSWDPPEN	14688	Db	15709	TIKDACREDVGHVVKLTNSAGEALETNLVLDKPGPPTGVKMDVETADTSITLSWGPP	15768
				QY	290	-----	289

Db	15769	KYDGGSSINNIYI	VEKRDSTTTWQI	VSATVARTTI	KACRLKTCGEYQ	FRIAAENRYGKST	15828
QY	290	-----	-----	-----	-----	-----	289
Db	15829	YLNSEPTVAQY	PFKVP	PGPGT	PVVTLS	SRDSMEVQWNEPIS	DGGSRVIGYHLERKERN
QY	290	-----	-----	-----	-----	-----	289
Db	15889	LWVLANKTPI	PQTKFTT	GLEGEVYEF	FRVSAENI	VGIGKPSVSECV	VARDCPPGRP
QY	290	-----	-----	-----	-----	-----	289
Db	15949	EALIVTRNSV	TLOWKPYTDG	SKITGI	YI	VEKKELPEGR	WMKASFTNIIDTHPEVTGLVE
QY	290	-----	-----	-----	-----	-----	289
Db	16009	DHRYEFRV	IARNAAGV	SEBSEST	GAITARDE	VDP	PRISMDPKYKDTIVVHAGESFKVDA
QY	290	-----	-----	-----	-----	-----	289
Db	16069	DIYGKPI	TIQWIKG	QELSN	TARLEIK	STD	FATSLSVKDAVRVDSGNVILKAKWAGER
QY	290	-----	-----	-----	-----	-----	289
Db	16129	SVTNVK	VLD	RP	GP	BGPVVISG	TAEKCTLAWKPLQDGGSDIINYIVERRETSRLVMT
QY	290	-----	-----	-----	-----	-----	289
Db	16189	VVDANV	QTL	SCVTK	LL	EGNEY	TRFIMANVKNYGVGBPLEBSEPVVAKNPVVPDAPKADEV
QY	290	-----	-----	-----	-----	-----	289
Db	16249	TTVT	KDSMI	VWER	PAS	DGSEIL	GYVLEKDEKIRWTRCHKLLIGELRLRTVGLIENH
QY	290	-----	-----	-----	-----	-----	289
Db	16309	DYEF	RVSA	ENAG	LS	BSP	SPSAYQKACDPIYKPGPNPKVIDITRSSVFLS
QY	290	-----	-----	-----	-----	-----	289
Db	16369	GCEI	QYI	VEK	CDVNV	GEW	TMCPTPTGINKNTIEVEKLLKHEYNFRICAINKAGVGEHA
QY	290	-----	-----	-----	-----	-----	289
Db	16429	DVPG	PII	VEK	LEAP	DIDL	LELKIINIRAGGSLRLFVPIKGRPTPEVKWGV
QY	290	-----	-----	-----	-----	-----	289
Db	16489	AAI	IVT	SS	FTSL	VLD	NVNRYS
QY	290	-----	-----	-----	-----	-----	289
Db	16549	TKDS	IVT	WE	PL	LDG	SKIKNYI
QY	290	-----	-----	-----	-----	-----	289
Db	16609	RVTA	ENEY	GIGL	PAQ	TAD	PIKVAEV
QY	290	-----	-----	-----	-----	-----	289
Db	16669	IVEM	QAKH	SEK	WSE	CAR	VKSQ
QY	290	-----	-----	-----	-----	-----	289
Db	16729	LVIE	P	V	K	P	A
QY	290	-----	-----	-----	-----	-----	289
Db	16789	TLSI	KETH	KDDG	QY	G	I
QY	300	-----	-----	-----	-----	-----	299
Db	16849	PPLY	TG	GC	CO	IT	NYI
QY	300	-----	-----	-----	-----	-----	299
Db	16909	SFA	LES	D	P	I	V
QY	300	-----	-----	-----	-----	-----	299
Db	16969	SIL	W	T	K	V	N
QY	300	-----	-----	-----	-----	-----	299
Db	17029	TPE	P	M	V	K	N
QY	300	-----	-----	-----	-----	-----	299
Db	17089	TED	Q	R	V	E	F
QY	300	-----	-----	-----	-----	-----	299
Db	17149	EAD	V	H	G	K	P
QY	300	-----	-----	-----	-----	-----	299
Db	17209	SKS	F	P	V	N	K
QY	300	-----	-----	-----	-----	-----	299
Db	17269	WTV	V	A	S	E	V
QY	300	-----	-----	-----	-----	-----	299
Db	17329	EVT	N	I	A	K	S
QY	300	-----	-----	-----	-----	-----	299
Db	17389	DH	E	F	R	V	S
QY	300	-----	-----	-----	-----	-----	299
Db	17449	DG	S	E	I	L	G
QY	300	-----	-----	-----	-----	-----	299
Db	17509	AT	S	P	G	T	V
QY	300	-----	-----	-----	-----	-----	299
Db	17569	TD	K	V	I	E	K
QY	300	-----	-----	-----	-----	-----	299
Db	17629	EY	R	K	S	A	F
QY	300	-----	-----	-----	-----	-----	299
Db	17689	YF	R	V	A	E	N
QY	300	-----	-----	-----	-----	-----	299
Db	17749	G	V	V	E	M	Q
QY	300	-----	-----	-----	-----	-----	299
Db	17809	KOL	T	Q	S	L	K
QY	300	-----	-----	-----	-----	-----	299
Db	17869	ST	V	L	H	I	K
QY	300	-----	-----	-----	-----	-----	299
Db	17929	WB	P	P	A	T	G
QY	300	-----	-----	-----	-----	-----	299

QY	312	-----	311	QY	327	-----	326
Db	17989	GKSAPLDSKAVIQYPPFKPGPGTFFVTSISKDQMLVQWHEPVNDGGTKIIGYHLEQXE	18048	Db	19069	RYKSSYSESSAVVAEYFPSPGPGTKVHVHATKSTMLVTWQVPVNDGSRVIGYHLEY	19128
QY	312	-----	311	QY	327	-----	326
Db	18049	KNSILWVKNLKTPIQDTKPTTCLDSGLEVPKVSANENVIGIKPSKVSECFVARDPCDP	18108	Db	19129	KERSILWSKANKILIAQTQVKVSGLDGLMYRVRVAENIAGIGKSKSEFVPARDPC	19188
QY	312	-----	319	QY	327	-----	326
Db	18109	PGRPEAIVTRNNVTLKMKPAYDGGKITGYIVEKKOLPDGRWMKASFNTVLETFVTS	18168	Db	19189	DPQOPEVTNITRKSLSKSPHYDGGAKITGYIVERRELDPDGRWLKCNVTNIQETYPE	19248
QY	320	-----	319	QY	327	-----	326
Db	18169	GLVEDORYEFRVIARNAAGNFSEPSDSSGAIYARDEIDAPNASLDPKYKDVIVHAGETF	18228	Db	19249	VTELTEDQRYEFRVFARNAADSVSEBSTGPIIVKDDVEPPRVMMVKFRDVIIVKAGE	19308
QY	320	-----	319	QY	327	-----	326
Db	18229	VLEADIRGKPIPDVVWSKDGKELEETAARMEIKSTIQKTTLVVKDCIRTDGGQYILKLSN	18288	Db	19309	VLKINADIAGRPLPVISWAKDGIIEBERARTEIISTDNHTLLTVKDCIRRDGTGVVLTK	19368
QY	320	-----	319	QY	327	-----	326
Db	18289	VGGTKSIPITVKVLDPRGSPGPKVTGVTAEKCYLAWNPLODGGANISHYIIKRETS	18348	Db	19369	NVAGTRSVAVNCKVLDKPPGAPGLEINGLTAEKCSLSWGRPOEDGADIDYVHKKRET	19428
QY	320	-----	322	QY	327	-----	326
Db	18349	RLSWTQVSTEVQALNYKVTKLLPGNEYIFRVMAVNYGIGEPLESPVTACNPYKPPGP	18408	Db	19429	SHLAWTICEGELQMTSCKVTLLKNGEYIFRVTVGNKYGVGEPLSEVAIKALDPFTVPSP	19488
QY	323	-----	326	QY	327	-----	326
Db	18409	STPEVSAITKDSMVVTWARPVDDGGTEIGYLEKRDKEGVRWTKCNKKTLDLRLRVTG	18468	Db	19489	PTSLEITSVTKESMTLCSWRPESDGGSEISGYIERREKNSLRWVRVNNKFPYVDLRVST	19548
QY	327	-----	326	QY	327	-----	326
Db	18469	LTEGHSYFRVAENAAGVGEPSVFRACDALYPPGPPSNPKVTDTSRSSVLSAWSK	18528	Db	19549	GLRECEYEYRVYAENAGLSLPSETSPILRAEDPVFLPSPSPKPIVDSGKTITIAWV	19608
QY	327	-----	326	QY	327	-----	326
Db	18529	PIYDGGAPVGYVVEKEAADWTCTPTGLQKQFTVKLKENTENFRICAINSEG	18588	Db	19609	KPLFDGGAPITGYTVVEYKKSDDTDWKTISIQSLRGTEYITISGLTTGAEYVFRVSKNVKA	19668
QY	327	-----	326	QY	327	-----	326
Db	18589	VGEPATLPGSVAQERIEPPEIELDADLRKVVLRASATLRLFTVTKGRPEPEVKWEAE	18648	Db	19669	SDPSDSDPQIAKEREEREPFLDIDSEMRKTLIVKAGASFTMTVPRGRPVNVLMSKPT	19728
QY	327	-----	326	QY	327	-----	330
Db	18649	GILTDRAQIEVTSFTMLVIDNVTRFDSGRYNLTLENNSGSKTAFVNVRLDPSAPVNL	18708	Db	19729	DLRTRAYVDTTDSRTSLTIENANRNDGKYTLTIQNVLSAASLTLVWKVLDTPGPPTNIT	19788
QY	327	-----	326	QY	331	-----	330
Db	18709	TIREVVKDSVTLSEPPIDGGAKITNYIVEKRETRKAVATITNNTKTTFRLENLQEG	18768	Db	19789	VQDVTKESAVLSWDVPENDGGAPVKNYHIEKREASKAWSVVTNNCNRLSYKVTLNQEGA	19848
QY	327	-----	326	QY	331	-----	330
Db	18769	CSYVFRVLASNEYIGLPAETTEPVKSEPPLPGRVTLVDVTRNTATIKWEKPESDGGS	18828	Db	19849	IYVFRVSGENERGVGIPAEKTEGVKITEKPSPEKLGVTISIKDSVSLTWLKEHDGSR	19908
QY	327	-----	326	QY	331	-----	332
Db	18829	KITGYVEMQTKSEKSTCTQVKTLEATISGLTAGEEYVFRVAANVEKGRSDPRQLGVP	18888	Db	19909	IVHYVVEALEKQKNWKCVAKSTHHVVVSGLSENSEYFFRVFAENQAGLSDPRELLLPV	19968
QY	327	-----	326	QY	333	-----	332
Db	18889	VIARDIEIKPSVELPFHTFNVKAREQLKIDVPFKGRPOATVNNRKDQTLKETTRVNSS	18948	Db	19969	LIKEQLEPPEIDMKNFPSSHVVYVRAGSNLKVDIPISGKPLPKVTLSRDGVPLKATMRFNT	20028
QY	327	-----	326	QY	333	-----	332
Db	18949	SKTWTSLSIKASKEDVGTIELCVSNSAGSITVPTIIVILDRPFGPPPIRIDEVCSDSIT	19008	Db	20029	EITAENLTINKESVTADAGRYEITAAANSSGTTKAFINIVLDRPGPPTGPPVISDITEE	20088
QY	327	-----	326	QY	333	-----	332
Db	19009	ISWNPPEYDGGCQISNYIVEKETTSTTWHIVSQAVARTSIKIVRLTTGSEYQFRVCAEN	19068	Db	20089	SVTLKWEPPKYDGGSOVTNYILLKRETSVAVWTEVSATVARTMVKMKLTTGEEYQFRIK	20148
				QY	333	-----	335



Db	20149	ANRFGISDHISACVTVKLPYTTTPGPPSTFWNTNTRBSITVGMHPVSNNGSVAVGXH	20208	Db	21229	EYQFRVRAENRIGVSOPLVSSIIVAKHQFRIQPGPGKPIYNYVTSOGMSLTWDAFYDGG	21288
QY	336	-----	335	QY	352	-----	351
Db	20209	LEMKDRNSILWQKANKLVIRTHFKVTTISAGLIYEFVRYAENAAAGVKPSHPVLAI	20268	Db	21289	SEVTGFHVEKKERNSTLWQKVNTSPISGREYRATGLVEGLDYQFRVYAENSAGLSSPSDP	21348
QY	336	-----	335	QY	352	-----	351
Db	20269	DACEPPNRVRIITDISKNSVLSWQPAFDGSGKITGYIVERRDLPGRWTKASFNTVET	20328	Db	21349	SKFTLAVSPVDPGTPDYIDVTRETITLKNPPLRDGGSKIYGVYSIEKQGNRWRCNF	21408
QY	336	-----	339	QY	352	-----	351
Db	20329	QFTISGLTQNSQYEFVRFARNAGVSNPSEVGPITCIDSYGGPVIDLPLEYTEVVKYR	20388	Db	21409	TDVSECQYTTVGLSPGDRYEPRIIARNAVGTISPSPQSSGIIMTRDENVPPIVEFGPEYF	21468
QY	340	-----	339	QY	352	-----	351
Db	20389	AGTSVKLRAGISGKPAPTIEWYKDKELQTNALVCVENTTDLASILI KDADRLNSCYEL	20448	Db	21469	DGLIIKSGBSLRIKALVQGRPVPRVTWFKDGVIEKRMMEITNVLGSTSLFVRDATRDH	21528
QY	340	-----	339	QY	352	-----	351
Db	20449	KLRNAMESATIRVQILDKPGPGGPIBFKVTVAEKITLLMRPPADDDGAKITHIVEK	20508	Db	21529	RGYTVTEAKNAGSASAKAEIKVKVQDTPGKVGPPIRFTNITGEKMTLMDAPLNDGCAPIT	21588
QY	340	-----	339	QY	352	-----	356
Db	20509	RETSRVVMSVSEHLEECITTTKIIKNEYIFRVRANVKYIGIBPLESDSVVAKNAFVT	20568	Db	21589	HYIIKRETSRLAWALIEDKCEAQSYTAIKLINGNEYQFRVSAVWKFGVRPLDSDPVVA	21648
QY	340	-----	339	QY	357	-----	356
Db	20569	PGPPGPIEVTKITKNSMTVWSRPIADGSDISGYFLEKRDKSLGMPKVKLETIRDTQ	20628	Db	21649	QIQYTVDPAPGIPESNITGNSITLTWARPESDGGSEIQOYILERREKSKTRVWKVSKR	21708
QY	340	-----	347	QY	357	-----	356
Db	20629	KVTGLTENSQYVRVCAVNAAGQGPSEPFYKAADPIDPPGPPAKIRIADSTKSSITL	20688	Db	21709	PISETRFKVTGLTEGNEYEFHVAENAAAGVPASGISRLIKCREPVNPPGPTVVKVTD	21768
QY	348	-----	347	QY	357	-----	356
Db	20689	GWSKPYVDCGSVAVTYVVEIROGEEBETTVSTKGEVRTTEYVYVSNLXPGVNYFRVSAV	20748	Db	21769	SKTIVSLEWSKVPDGGMEIIGYIEMCKTDLGWHKVNABACVTRYTVTDLQAGEEYK	21828
QY	348	-----	347	QY	357	-----	363
Db	20749	NCAGQGEPIEMNEPVQAKDILEAPEIDLVALRTSVIAKAGEDVQVLIIPFKGRPPPTVTW	20808	Db	21829	FRVSAINGAGKDSCEVTGTIKAVDRLTAPELDIDANFQTHVVRAGASIRLFIAYQGRP	21888
QY	348	-----	351	QY	364	-----	363
Db	20809	RKDEKNLGSARYSIENTDSSLLTIPQVTRNDTGKVIILTENGVCPEKSTTVSVKVLDT	20868	Db	21889	TPTAVMSKPSNLSLRADIIHTDTSFSLTVENCNRNDAGKYTLTVENSSGSKSITFTVKV	21948
QY	352	-----	351	QY	364	-----	363
Db	20869	PAACQKQVHVSRGTVTLLMDPPLIDGGSPIINYVIEKRDATKRTWSVSVSHKCSSTPK	20928	Db	21949	LDTGPPGPIITPKDVTGRSATLMDAPLLDGGARIHHYVVEKREASRRSQVISEKCTRQ	22008
QY	352	-----	351	QY	364	-----	363
Db	20929	LIDLSEKTPFFRVLAEINEIGIEPCETTEPVKAAEVPAPIRDLSMKDSTKTSVILSWTK	20988	Db	22009	IFKVNDLAEGVPYFRVSAVNEVGVGEPEYEMPEPIVATQAPPRRLDVVDTSKSAVLA	22068
QY	352	-----	351	QY	364	-----	363
Db	20989	PDFDGGSVITEYVVERKGEQWTHAGISKTCEIEVSQLEKQSVLEFRVFAKNEKGLSD	21048	Db	22069	WLKPDHGGSRITGYLLEMRQKGSOLWVEAGHTKQLTFTVERLVEKTEYEYFRVAKANDAG	22128
QY	352	-----	351	QY	364	-----	363
Db	21049	PVTIGPITVKELIITPEVDLSIDPGAQVTVRIGHNVHLELPYKGPSPISMLKGLPLK	21108	Db	22129	YSEPREAFSSVIIKBPQIEPTADLTGITNQLITCKAGSPFTIDVPISGRPAPKVTWKLSE	22188
QY	352	-----	351	QY	364	-----	363
Db	21109	ESEFVRFSKTENKITLSIKNAKEHGKVTVIDLNAVCRIVPITVITLGPSPKPGPIR	21168	Db	22189	MLKETDRVSIITTKDRTTLTVKDSMRGDSGRYFLTLENTAGVKTFSTVTVVIGRPGPVT	22248
QY	352	-----	351	QY	364	-----	363
Db	21169	FDEIKADSVILSWDPEDNGGGEITCYSIKRETSQTNMKWCVSSVARTTPKVPNLVKDA	21228	Db	22249	GP1EVSSVSAESCVLWSGEPKGGGTEITNIVEKRESGTTAQLVNLSSVKTQIKVTHL	22308
QY	352	-----	351	QY	364	-----	370
Db	22309	TKYMEYSFRVSSNRFVSGKPLESAPIIAEHPFVPSAPTRPEVYHVHVSANAMSIRWEEPY	22368	QY	364	-----	370

QY	371	-----	370	QY	382	-----	381
Db	22369	HDGSKIIGYWEKBERNTILWVKENKVPCLNCKYKVTGLVEGLEYQRTYALNAAGVSK	22428	Db	23449	DGMTLWYPPEDDGGSOVTGYIVERKEVRADRWVRVNVKVPVTWTRYSTGLTEGLEVEHR	23508
QY	371	-----	370	QY	382	-----	381
Db	22429	ASEASRPIMAQNVPDAPGRPEVTDVTRSTVLSWAPAYDGGSKVVGYYIERKPVSEVD	22488	Db	23509	VTAINARGSGKSRPSKPIVAMDPIAPPGKPQNPRTDTRTSTVSLAWSVPDEGGSKVT	23568
QY	371	-----	370	QY	382	-----	381
Db	22489	GRWLKCNVTIVSDNFTVTALSEGDTYPRVLAKNAAGVISKSBSGTGPTCRDEYAPPK	22548	Db	23569	GYLIEQMVDQHEWTKCNTTPTKIREYTLTHLPOGAERYFRVLACNAGGPGPAEVPCTV	23628
QY	371	-----	370	QY	382	-----	389
Db	22549	AELDARLHGLVTIRAGSOLVDAAVGGKPEPKIITWTKGDKELDLCCKVSLQYTKRATA	22608	Db	23629	KVTEMLEYPDYELDERYQEGIFVRQGGVIRLTIPIKGPPPICKWTKGQDISKRAMIAT	23688
QY	371	-----	370	QY	390	-----	392
Db	22609	VIKFCDRSDSGKTLTVKNASGTKAVSMVKVLDSPGCGKLTVRVTOEKTALWSLPQ	22668	Db	23689	SETHTELVIKEADRGDGTVDLVLENKCGKANVYIKURVIGSPNSPBGPLEYDDIOVRSV	23748
QY	371	-----	370	QY	393	-----	392
Db	22669	EDGGABIITHYIVERRETSLNWWIVEGECPTLSYVVTRLIKNNEYIFRVRVKNYGPVGP	22728	Db	23749	RVSMRPPADGGADILGYILLERREVPKAAWYTIDSRVGTSLVVKLKENVEYHFRVSAE	23808
QY	371	-----	370	QY	393	-----	392
Db	22729	VESEPIVARNSTPIPSPGPIPEVGTGKEHIIQWTKPESDGNELSNVLDKREKESLR	22788	Db	23809	NQFGISKPLKSEEPVTPKTLNPPPEPSNPPEVLDTKSSVLSWSRPKDDGGSRVTGY	23868
QY	371	-----	370	QY	393	-----	406
Db	22789	WTRVNDYVYDTRLKVTSILMEGCDYQFRVAVNAAGNSEPERSNFISCRPSYTPGPP	22848	Db	23869	IERTSTDKVRHNKQITTTMYTVTGLVPAEYQFRIIAQNDVGLSETSPASEPVVCK	23928
QY	371	-----	370	QY	407	-----	409
Db	22849	SAPRVVDTTKHSISLAWTKPMYDGTDIVGYVLEMQEKDQWRYVHTNATIRNTEFTVP	22908	Db	23929	DPFDKPSQGELEILLSISKDSVTLQWEKPECDGKBEILGYVWEYRQSGSAWKSNKERI	23988
QY	371	-----	370	QY	410	-----	409
Db	22909	DLKMGQKISPRVAANVKGMSYESIAEIEPVERIEIPDLLELADDLKKTVTIRAGASLR	22968	Db	23989	KDKQFTIGLLEATEYEFVRFAENETGLSRPRRTAMSIKTLTSGEAPGIRKEMKDVTYK	24048
QY	371	-----	370	QY	410	-----	409
Db	22969	LMVSGRPPPVITWSQKIDLASRAIDTTESYSLLIVDKVNRVDAGKYTIEAENOSGK	23028	Db	24049	LGEAAQLSCQIVGRPLPDIKWYRFGKELIQSRKYKMSDGRHTLTVMTEQDEGVYTC	24108
QY	371	-----	370	QY	410	-----	409
Db	23029	KSATVLVKVYDTGPCPSVKVKEVSRDSVTITWEIPTIDGAPINNVIVEKREAMRAFK	23088	Db	24109	IATNEVGEVETSKLLQATPFQHPGYPLKEKYGAVGSTLRHLHVMYIGRPVPAMTWFG	24168
QY	371	-----	374	QY	410	-----	409
Db	23089	TVTTKCKTLRISGLVEGTMHYFRVLNENYIGIGBPCETSDAVLSEVPLVPAKLEVVD	23148	Db	24169	QKLLONSENITIENTEHTHYLVMKNVQRKTHAGKYKQLSNVFGTVDAILDVEIQDKPK	24228
QY	375	-----	374	QY	410	-----	409
Db	23149	VTKSTVTLAWKPLYDGSRLTGYLBACKAGTERMMKVVLKPTVLEHTVTSLSNEGEQY	23208	Db	24229	PTGPIVIEALLKNSAVISWKPPADGGSWITNYYVVEKCEAKEGAEWQLVSSAISVTTCRI	24288
QY	375	-----	381	QY	410	-----	417
Db	23209	LFRIRANEKGVSEPRETVTAATVQDLRVLPITDLSMPQKTHVPAGRPVELVPIAGR	23268	Db	24289	VNLTENAGYFRVSAQNTFGISDPLEVSSVLIKSPFEKPGAPGKPTTAVTKDSCVNAV	24348
QY	382	-----	381	QY	418	-----	417
Db	23269	PPPAASWFFAGSKLRESERVVETHTKVAKLTIRETTIRDTGEYTLBLKNVTGTTSETIK	23328	Db	24349	KPPASDGGAKIRNYILEKKEKQNKWISVTTBEIRETVFSVKNLIEGLEBYEPRVKCNELG	24408
QY	382	-----	381	QY	418	-----	417
Db	23329	VIIIDKPGPTGPIKIDEIDATISITISWPELPGAPLSGVYVEQORDAHRPGLPVSES	23388	Db	24409	GESEWSEISEPITPKSDVPDPIQAPHKBEELNLRVYQSNATLIVCKVTGHPKPIVKWYROG	24468
QY	382	-----	381	QY	418	-----	417
Db	23389	VTRSTFKFTRLTGEGYVFRVAATNRFGISGLQSEVIECKRSSIRIPGPPETLQIFDVSR	23448	Db	24469	KEIIADGLKYIQEFKGYHQLIIASVTDDATVYQVRATNOGGSVSGTASLEVEVPKI	24528
				QY	418	-----	417

Db	24529	HLPKTLGMAVHALRGEVVISIKIPFSGKPDVPITWKGQDLIDNNGHYQVIVTRSFSL	24588	Db	25609	KRTEERLLEBELELGFSPSPSRPPHPELSSLYSSPQAHVKVEETRONFRYSTYHI	25668
QY	418	-----	417	QY	437	-----	436
Db	24589	VFPNGVERKDAGPVVYVCAKRFIDQKTVELDVADVPDPGRGVKVS DASRDSVNLWTWEP	24648	Db	25669	PTKABASTYAELERRERHAQAYRQPKQRORIMAREDEBELLRPVTTTQHLSEYKSELDFM	25728
QY	418	-----	417	QY	437	-----	443
Db	24649	ASDGSKITNYIVEKCAATTAERLVRQARETRYVINLFGKTSYQFRVIAENKFGLSKP	24708	Db	25729	SKEKSRKRSRQREVTEITEIEEVEYISKHAQRESSASRLRRRSLSPYIELMRP	25788
QY	418	-----	417	QY	444	-----	445
Db	24709	SEPSEPTITKEDKTRAMVDBEVEDRETVSMTKASHSTKELYKMYIAEDLGRBFGIV	24768	Db	25789	VSELIRSRQPAABEYEDDTERRSPTRPRSPSPVSSERSLSRFSARGPDI FSRYES	25848
QY	418	-----	417	QY	446	-----	445
Db	24769	HRCVETSSKTYMAKFVKVGTDOVLVKEKISILNIAHRNHLHSHESMEELWIFE	24828	Db	25849	MKAALKTKTSEKYEVLQQPPTLDHAPRIITLRMRSHRVPQCONTRFILNVQSKPTAEV	25908
QY	418	-----	417	QY	446	-----	445
Db	24829	FISGLDIFERINTSAFELNEREIVSVHVCEALQFLSHGNIGHFDIRPENIIYQTRSS	24888	Db	25909	KWYHNGVELQESSKIHYTNTSGVLTLLEILDCHTDDSGTYRACVCTNYKGEASDYATLDVTG	25968
QY	418	-----	418	QY	446	-----	445
Db	24889	TIKIEFGOARQLKPGDNFRLFTAPEYYAPEVHOHDVSVTATDMMSLGLTVYVLLSGIN	24948	Db	25969	GDYTYVASORRDEEVPRSVFPBELTRTEYAVVPSFKTSEMEASSSVREVKSQMTETRESL	26028
QY	419	-----	425	QY	446	-----	445
Db	24949	PFLAETNQOIENIMNAEYTFDEEAFKEISIEAMDFVDRLLVKEKSRMTASEALQHPWL	25008	Db	26029	SSVEHSASAEMKSAALEKSEKSTTRKIITTLAARILTKPRSMTVYGESARFSCDTD	26088
QY	426	-----	425	QY	446	-----	445
Db	25009	KQKIERVSTKVIRTLKHRRYHTLKKDLNMVVSAAISCGGAIRSQKGVSAKVAVASI	25068	Db	26089	GBPVPVTWLRKGQVLSTARSQVTTTKYKSTFEISSVQASDEGNSVSVVSENSEKQAE	26148
QY	426	-----	433	QY	446	-----	452
Db	25069	EIGPVSQIMHAVGEEGHVKYCKIENYDQSTQVTWYFGRQLENSEKYEITYEDGVAI	25128	Db	26149	FTLTQKARVTEKAVTSPPRVKSPEPRVKSPEAVKSPKRVKSPSPKAVSTETKPTP	26208
QY	434	-----	433	QY	453	-----	452
Db	25129	LYVKDITKLDGTYRCVNDYGEDSSYAELEFVKGVREVDYCYCRTMKIKERTDTMRL	25188	Db	26209	REKVQHPVSAAPPKITQPLKAEASKEIAKLCVVBESSVLRAKEVTWYKDGKCLKENGHFQ	26268
QY	434	-----	433	QY	453	-----	458
Db	25189	LERPPEFTLPLYNKTA YGVENRFGVTITVHPBPHVTWYKSGQIKPGDNDKKYTPESDK	25248	Db	26269	PHYSADGTYELKINNLTESDQGEYVCEISGEGGTSKTNLQFMGQAFKSIHEKYSKISETK	26328
QY	434	-----	433	QY	459	-----	461
Db	25249	GLYQLTINSVTTDDAEYTVVARNKYGEDSCAKLTVTLHPPTDSTLRPMPKRLLANAB	25308	Db	26329	KSDQKTTESTVTRKTEPKAPBPISSKPVI VTGLQDTTVSSDSVAKPAVKATGEPRTAIW	26388
QY	434	-----	433	QY	462	-----	461
Db	25309	CQEGQVCFEIRVSGIPPPTLKWEKGQPLSLGNPIEIIHEGLDYALHIRTLPEDTGY	25368	Db	26389	TKDGKAITOGKVKLSEDKGPFLEIHKHTDTS DGLYTCVKNAGSVSSSCKLTIKAIK	26448
QY	434	-----	433	QY	462	-----	461
Db	25369	YRVTAINTAGSTSCQAHLQVERLRKYKQBFKSEHERHVQKQIDKTLRMAELSGTESV	25428	Db	26449	DTEAQKVSTQKTSEITPQKAVVQBEEISQALRSEBIKMEAKSQBKALKEBASKVLIS	26508
QY	434	-----	433	QY	462	-----	461
Db	25429	PLTVAKEALREAAVLYKPAVSTKTVKGBFLBIEBKBERKLRMPYDVPEPRKYQTITI	25488	Db	26509	EEVKSAATSLEKSI VHEEITKTSQASEVRTHAEIKAFSTQMSINEGORLVLKANIAGA	26568
QY	434	-----	433	QY	462	-----	462
Db	25489	EEDQRIKQFVMSDMKWKYKIRDOYEMPGKLD RVQKRPKRIRLSRWEQFYVWPLPRI TD	25548	Db	26569	TDVKVNLNGVELTNSSEYRYGVSGSDQTLTIKASHRDESGILTCISKTKEGIVKCOYDLT	26628
QY	434	-----	433	QY	463	-----	465
Db	25549	QYRPKWIRPKLSQDDLEIVRPARRTPSPDYDFYPRRRLSGDISDEHLLLPIDDYLAM	25608	Db	26629	LSKELSDAPAFISQPRSONINEGQNVLF TCEISGSPSPBIENFK 26672	
QY	434	-----	436				

ID XX ABM67171 standard; protein; 16368 AA.  
AC XX ABM67171;  
XX XX  
DT 20-NOV-2003 (first entry)  
DE XX  
DE XX Photorhabdus luminescens protein sequence #268.  
XX XX  
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
KW detection; food; gene expression; plant; animal; microorganism; toxin;  
KW antibiotic; biopesticide; virulence factor; disease model; plague;  
KW whooping cough.  
XX XX  
XX Photorhabdus luminescens.  
OS XX  
XX WO200294867-A2.  
XX XX  
XX 28-NOV-2002.  
XX XX  
XX 07-FEB-2002; 2002WO-IB003040.  
XX PF  
XX 07-FEB-2001; 2001FR-00001659.  
XX PR  
XX (INSP ) INST PASTEUR.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX XX  
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
PI Buchrieser C;  
XX XX  
XX WPI; 2003-148459/14.  
XX XX  
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX XX  
XX Claim 2; SEQ ID NO 268; 1205pp; French.  
XX XX  
CC The invention relates to the isolation of genes and their encoded  
CC proteins from Photorhabdus luminescens. The isolated sequences are  
CC sources of probes and primers for detecting the genome of P. luminescens  
CC and related species; to study polymorphisms; for gene analysis and for  
CC detection/amplification of the genes. Antibodies (Ab) raised against the  
CC polypeptides encoded by the genes are used for detection/identification  
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
CC carry a gene-containing vector are used to select compounds that  
CC modulate, regulate, induce or inhibit expression of the genes in plants,  
CC animals or microorganisms other than P. luminescens and are able to alter  
CC response or sensitivity to toxins and antibiotics produced by P.  
CC luminescens. Cells transformed to express the genes are useful for  
CC recombinant production of the proteins, particularly toxins and  
CC antibacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically (to treat microbial infection by bacteria or fungi that  
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luminescens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luminescens proteins  
XX XX  
SQ Sequence 16368 AA;  
Query Match 23.7%; Score 581; DB 6; Length 16368;  
Best Local Similarity 1.6%; Pred. No. 0.00032;  
Matches 246; Conservative 88; Mismatches 131; Indels 14861; Gaps 84;  
QY 1 MGRKE-----5  
DB 996 LGENDQQVKIRGRIEPIEIRLMEHPAVSEALVIALNDGQDKELVAVYMAEADDQLVN 1055  
QY 6 -----5  
DB 1056 SLRAYLSAILPDYMWPTAFVRLDTPLTPNGKLDRRALPAGEEAFARQAYAAPQGETES 1115  
QY 6 -----MMVRDV-----11

Db 1116 LLSAVMGELIGIEQVSRHDSFFALGCHSLLAIVRMVERLRNAGLTLAVRDLFOSPVLSIFA 1175  
QY 12 -----PKMFVLISIS-----21  
Db 1176 QTVGSOAVIVPANIMTPAITALTPEMLPLIDLTQSEIDHIVAQVPGGVANIQDIYGLSP 1235  
QY 22 -----FLLV-----26  
Db 1236 LQDGILPHLLANKGDPYLFVSKWVFADRALDLSYLVAVQVVRNHDILRTAFVMOGLSA 1295  
QY 27 -----26  
Db 1296 PVQVWRQAQSVTELTLPDVGFPVRDQAQYFDLHHRDLNQAPLLHFGVAQEEDGRW 1355  
QY 27 -----26  
Db 1356 SVLQLLHLIGDHTTLEVMNGEIQAYLAGQEGNLPVPIPPRLNVAQARLGVSQAEHTRFF 1415  
QY 27 -----26  
Db 1416 TEMLAEDQPTLPFGLMEVHRDGSQVTSYRMLPATLNDRLRSQAORLGVSLATLCHLAW 1475  
QY 27 -----FIN-----29  
Db 1476 AQLSHTSGQEKVVGTVLFGRMVQGGADSGMGLFINTLPLRLDINDASVDQSVQAVHT 1535  
QY 30 -----29  
Db 1536 RLAEILDHEHMSLALAQCRSRVPGEIPLFSALLNRYHNDQTFPENKEVPGIEFIGAERT 1595  
QY 30 -----CKWMSKAL-----37  
Db 1596 NYPFVLSVEDGGSTGLGTAQVQVFPFESERVCGYMQQALLESIAEALEQRPETPTVTLNLP 1655  
QY 38 -----37  
Db 1656 AAEKKLLLETWNATATETSYDPHSCIHRLFEQWAKKSPDATALVVEEQILSYANLNACAN 1715  
QY 38 -----37  
Db 1716 RLHQIALGVTPDQVRVAVCVSRPAMVAGVAVLAKAGAVPLDDPAPPSARLAHLTDA 1775  
QY 38 -----37  
Db 1776 APANVLADKAGKVALGEEALTGLTVLDNPLPNQVDSNPQISGLTSRHLAVVITSGSTG 1835  
QY 38 -----37  
Db 1836 VPKGWMIEHRTNPNFLYWAQAFETEIREVLFSTSMNFDLSIFECPVLSOGATIHIVE 1895  
QY 38 -----37  
Db 1896 DALSLMQHALPVTLINSVPSAMKPLLQALMASVHTVNLAGEPLKGLIEQIFETQIQ 1955  
QY 38 -----37  
Db 1956 RLCNLYGPSETTYSANLPIQGRDRIIESIGRIPIANTRYLLDENGQPVPLGVGVEIYIG 2015  
QY 38 -----YNRP-----41  
Db 2016 GAGVARGYFNRPDLTAERFLIDPFSDVPDARMYRTGDLARYFPDGNLEFLGRNDQVKIR 2075  
QY 42 -----41  
Db 2076 GFRIEPEGEARLMEHPAVSEAVILALDDDDQDKRLVAVYVAEADVLINRLHAHLASVLP 2135  
QY 42 -----WRGLV-46  
Db 2136 DYMPAAFCVCLDAPPLTPNGKLDRRALPAGGEDFARQIYAAPSGEMETTLAIRMRELUG 2195  
QY 47 -----LSKIG-----51

Db	2196	VERISRYDNFPALGCHSLLAVRMNRI	AALGIELPLSTLFTFPSLIAPAEVMRTQFDESG	2255	QY	64	-----	63		
QY	52	-----	-----	51	Db	3336	DHIEHQPGGVANIQDIYALS	PLQDGIPLFHLLANKGDPYLLVSQMI	FADRSLLDYLAA	3395
Db	2256	AALPAIVPLSREDHQPLSFAQQRLWFLAQ	NGVSEYTHIPLALRLGRGLDITAWQALD	2315	QY	64	-----	63		
QY	52	-----	-----	58	Db	3396	VOQVNRHDIKTAFWQGLSAPVQVVRQAQ	LSVTETLDPVDGVPVRDQLAQHFDPRH	3455	
Db	2316	TLFARHEALRSVFVAVDGPQVELLPASS	LLMKYDLRTPDVPDQLAFLSAQESAPF	2375	QY	64	-----	63		
QY	59	-----	-----	58	Db	3456	RLDLNAQPLLFHFAIAQEDDGRWSVLQ	LLHLLIGDHTTLEVMNGEVQAYLAGOBES	PAPA	3515
Db	2376	DLARGPLIRSSVLQADDEHVFLITQH	HIVSDGWSLILKSELEALYSAHLNQDPLP	2435	QY	64	-----	63		
QY	59	-----	-----	58	Db	3516	PFRLVAQARLGVSAEHTRFTDMLAEV	EPTLPFGLTEVYRDGSQVTEFHRMLPATLN	3575	
Db	2436	LTIQPDYAAQKOWLSAERIQSOSDY	WRARLVDA PVLLDLPTDRPRPQSFAGALLPV	2495	QY	64	-----	72		
QY	59	-----	-----	58	Db	3576	DLRSQARLDSLATLCHLAWAQLVSR	TSGREVVFGTVLFRMQAGOGADSGMGLFIN	3635	
Db	2496	NLDAELTQSLKRLSEQGVTLFWTL	LAAWTAVLSRLSGQEDLVVGTSPSAGRSRQEVESLI	2555	QY	73	-----	77		
QY	59	-----	-----	58	Db	3636	TLPLRLDMDETSVQDSVRLAHTRLA	GLEHEHSLALAQRCSGVLGEVPLFSALLNYRHS	3695	
Db	2556	GFVNTLALRMDLSGAPNUTELLARV	QTALAAQEHQDLPFQVQVIVQPPRLAHTPLP	2615	QY	78	KQPV	-----	81	
QY	59	-----	-----	58	Db	3696	TQPVTTDAIISGIELSEOBERTNYP	FLSVEDFGHALGLTQIVQIDPERICGYMQAL	3755	
Db	2616	QVIFAWQNNENIEWRLPGLAVSLADQ	PIDIAKFDLSLSLSEVDGRVGVLYNATALPDQA	2675	QY	82	-----	88		
QY	59	-----	-----	63	Db	3756	ESLVEALQEPPTPIRALNILPAEKR	LLELTWNATQAPYPOCCVHQHLEQOAGNTPDA	3815	
Db	2676	TIERQVGYLTVLRGMATNPQPVGKI	DLTVTERKLLLETWNATKTSYPDHGCHRLFE	2735	QY	89	-----	88		
QY	64	-----	-----	63	Db	3816	IALVYGEHLTSAELNARANWLAWL	IGOGIOPDDRVAVQLERSIELVVAQLAILKAGAV	3875	
Db	2736	QOAKTPDAIALVYKILSYAELNARAN	LAHQILITLGAPEQRVAICVASSPARIVGL	2795	QY	89	-----	95		
QY	64	-----	-----	63	Db	3876	YVPIDRPVDERKHWLISDCAAKLLT	DI PVDLADKIGAISGEDYENPDLPRISTELAYI	3935	
Db	2796	LAVLKAGAVVPLDPAYPGERLHI	LTDAAPIVLADNTGRAALGKVLAAALTVDLPNSL	2855	QY	96	L	-----	99	
QY	64	-----	-----	63	Db	3936	MYTSGSTGPKGVLPHRVARBLVN	NGVAETGDDRVAFGANPSPDASTFEVWAPLLNG	3995	
Db	2856	PQPDSPNPQVPLTPHLLAVIYVTS	GTGIPKGVWVEHCGLVNLVQEKIVQFDIHPGSRM	2915	QY	100	-----	109		
QY	64	-----	-----	63	Db	3996	GTLLVIDHATVLTPEPALALQTYR	ITVLTWLSVGLFNRLVAELS PVLQPKILLIVGSDVL	4055	
Db	2916	LQFASFGDAGVWETMMALCSGATL	AI PADTVRQEBPRYLWHYLEBRAITHACLTPALLRE	2975	QY	110	-----	109		
QY	64	-----	-----	63	Db	4056	DPHVMAQVLRDNPQQLLSAYG	SPSEGTFTTTCITALPSVVARIPGRPIANARIYLLD	4115	
Db	2976	GTDLPEMTIRPTLILGGEAPSTLL	QALCERATVFNAYGTEITVCAATWRCPSDYTEGV	3035	QY	110	-----	109		
QY	64	-----	-----	63	Db	4116	TYGQPVPLGAIGEIVYGGDVAC	GYNLNRPDQTAERFLTPFSDQPDARMYRTGDLARYLP	4175	
Db	3036	IAIGRPANTQIYLLNTDQOPVPLG	AVGELYIGGIVARGYLNRPDLTAERFLADPFSDK	3095	QY	110	-----	109		
QY	64	-----	-----	63	Db	4176	DGNLEFLGRNDQOVKIRGFR	IEPGEVEARLLEHPAVHEAVLVVDGQDKRLVAVYVAEA	4235	
Db	3096	PDARLYRTGDLARYLPDGNLE	FLGRNDQOVKIRGFRIEPGEVEARLLEHPAVHEAVLVVDGQDKRLVAVYVAEA	3155	QY	110	-----	109		
QY	64	-----	-----	63	Db	4236	DEMLVNLRLDLSAVLPDYVMPAA	FVCLDAFPLTPNGKLDRRALPVPGBEDFARQIYAVP	4295	
Db	3156	GDQDKELVAYVAEANEELVNLH	RLTHLSAILPDYVMPAA FVRLDAFPLTPNGKLDRRAL	3215	QY	110	-----	109		
QY	64	-----	-----	63	Db	4296	SGEMETTLAAIWRELLGVERIS	RYDNFFALGCHSLLAVRMNIRIALGIELPLSTLFTFP	4355	
Db	3216	PAPGAFAFARQVYAAPQGETETL	MAAIWSELLGIEQVSRHDSFFALGCHSLLAVRMVREL	3275	QY	110	-----	109		
QY	64	-----	-----	63	Db	4356	TLMAFAVMSARFDESGEALPAI	IPLSREGQLPLSFAQQRLWLLAQFEGVSETHIPLAL	4415	
Db	3276	RNAGLTLAVRDLFQSPVLSFEAQ	TGQSPVVVVVAVANVITPALTTLTPEMPLPIDLTQSEI	3335						

QY 110 ----- 109  
Db 4416 RLRLGDIATWQALDTLFAHEALRSVFVAVDGPQVBLLPAALGWPMMKYDLRNPDV 4475  
QY 110 ----- 109  
Db 4476 DAQLASLSQEAQTPPDLARGPLIRGALVOLANDEHVFLTLHHIIFDGWSVSLMCELN 4535  
QY 110 ----- 109  
Db 4536 ALYTAFLTGQDPPLSLTIQYDPDYAAWQWLSABRIQFQSDVWRARLVDAFVLLDLPTD 4595  
QY 110 ----- 109  
Db 4596 RPRPRQSFTGNILPVSLDAELTQSLKRLSEQGVTLFMTLLAAWATVLSRSGQBDLVI 4655  
QY 110 ----- 109  
Db 4656 GTPSAGRSQEVESLIGFFVNTLALRMDLSGAPNVUTELLAHVRQTALAAQEHQDLPFEQV 4715  
QY 110 ----- 112  
Db 4716 VEIVQPLRLAHTPLFOVIFAWQSNENAEWQLPGLTVLPADQLIDIAKFEDLELSLSEVDG 4775  
QY 113 ----- 117  
Db 4776 RIVGYLNVATALFDQATIERQVGYLHTVLRAMVANPQPVGKIDILTVAERKLLLETWSM 4835  
QY 118 ----- 117  
Db 4836 TETSYPDHGCIHRLFQWAEKNPDATALMYEQAISYASLNTSANRLAHQLITLGVVPDQ 4895  
QY 118 ----- 120  
Db 4896 RVAICVARSPAMIVGLLAVLKAGGAVPLDPAYSGERLGHILVDKPAILLADNTGRAAL 4955  
QY 121 ----- 120  
Db 4956 GBKALADLIMLPNILFDQADSNPLIPELTERLAYVVTYTSGSTGTPKGMVMEHGVNVL 5015  
QY 121 ----- 120  
Db 5016 ALAQITRFSVDETSRILQFASGFDASVSEIMTALSGGACLVIPTDIIRODPRLMSYLE 5075  
QY 121 ----- 120  
Db 5076 KOAVTHAFLPPALFREESDLPVITIKPTLIFAGEAPGSTLFRALCDRVNLFNDYGPTEIT 5135  
QY 121 ----- 120  
Db 5136 VCATSWSCPSDYTDWVPIGRPTTNRVYLLDTYQFVPLGTVGELYIGGVGVARGYLNY 5195  
QY 121 ----- 123  
Db 5196 PELTAERFLTDPFSNEPDARLYRTGDLARYLPDGNLI FVGRNDQVKIRGFRIEPOEIEA 5255  
QY 124 ----- 123  
Db 5256 RLMEHPAVSGALVLSGDQDKELVAYVVAEPDGLTASLRYLSAILPDYVMPAAFVRL 5315  
QY 124 ----- 123  
Db 5316 NTFPLTPNGKLORRALPAGADAFARQVYAAPOGETEILLATIWSBLLGIEQISRHDSFF 5375  
QY 124 ----- 123  
Db 5376 ALGGHSLAVRMIERLURNAGLTLAVHDLFQSPVLSAPAEVQSQAVVVPENVITPATTA 5435  
QY 124 ----- 123  
Db 5436 LTPEMPLPLIELTQTEIDYIVQQMPGGIVNIQDIYALSPLODGLTFHLLTNEGDPLYLVS 5495  
QY 124 ----- 123

Db 5496 QMAFADRSLLSYLVAVQQVNNRHTLRTAFIWOGLSAPVQVWVRQAPLSVVLEILD PAD 5555  
QY 124 --VYSQ----- 127  
Db 5556 GPVYEQLAQRFNPROHRLDLGOAPLLRFVIAQEMDGRWIALQOHHLLIGDHTTLEVMNRE 5615  
QY 128 ----- 127  
Db 5616 VQAYLTGBESLPVPAPFNLVAQVRLGVNQAEHTRFTFMDLAEVDQPTLPFLGTEVHRD 5675  
QY 128 ----- 127  
Db 5676 GSQMTQSYRMLPAIILNNRLRSQARLGVSLAALCHLAWAQLSRTSGQKVVGTVLFOR 5735  
QY 128 ----- 127  
Db 5736 MQTGQTDSSMGLFINTLPLRLDMDDTPVRGRVQAAHARLAELLEHEHASLAWAQRCSGI 5795  
QY 128 ----- 136  
Db 5796 QGEVPLFSALLNYRHSALSATSNEIMNGVEFLGAQERTNYPTLSVEDEFGNALGLTTQIV 5855  
QY 137 ----- 136  
Db 5856 QPFDPERVCGYMQALESIAETLELAPETFVRTLEILPEARTLLKTNWETEIAYSDPL 5915  
QY 137 ----- 136  
Db 5916 CIHQLEQOQVEKNPDATALVYEEQYSYAELNTRANLHAHQIALGIVPDQORVAICVTRS 5975  
QY 137 ----- 136  
Db 5976 PTMIVGLLAVLKAGGYVPLAPAYPGERLAHILTDAAPAILLADNVGCAALGEEALSGLT 6035  
QY 137 ----- 136  
Db 6036 VLDPNLTPKPSNPQVTALTQAHLAYVIYTSGSTGTPKGMVMEHRLNLVRDKIAQPD 6095  
QY 137 ----- 136  
Db 6096 IHSDSRILOFASLSPDASVWEIMMALGSGACLVIAVDIVRQDPLRLWHYLEQQVVTACL 6155  
QY 137 ----- 136  
Db 6156 TPALLRDGDLPALATITPTVILGGEAPSAAALFQTLCCRRAALFNAYGPTBITVCAATWRCP 6215  
QY 137 ----- 136  
Db 6216 PDYTDTLVPIGHPTANTQIYLLNSDQVPVLGAVGELYVGGAGVARGYLNRRBELTAERFL 6275  
QY 137 -----FR----- 138  
Db 6276 ADFFSEMPGARMYRTGDLARYLPDGNLVFVGRNDQQIKIRGFRIEGETRILEYPAVR 6335  
QY 139 ----- 138  
Db 6336 EAVVLARGEHDKRLVAVWVAEBENDELANSRLTHLSTILPDYMPAAFVRLDALPLTPNG 6395  
QY 139 ----- 138  
Db 6396 KLDRLALPADDEAFSRQIYEAPQGETETALAAIWRELLIGIELVGRYDSFFALGGHSLLA 6455  
QY 139 ----- 138  
Db 6456 VRMINMAALGIELPLTTLFKSPTLADFAEVMRAQLGEQNNSVSAMLPIISREGALPLSPA 6515  
QY 139 ----- 138  
Db 6516 QORLMLLTQFEGVSQTYHVPWALRLHGQDIAAQQOALNRLFARHEALRSIFISEDQPO 6575  
QY 139 ----- 138

Db	6576	VELLPABELGPMKHDRLKVPFDADKRLERLCVQEAEPFDLAGGPIRAGLIQRADDDYV	6635
Qy	139	-----	138
Db	6636	FLLTQHIIYDGSAGVLMHLSALYTAFSMGLPDPPLPLAIQYPDYAAQRCWLSABRL	6695
Qy	139	-----	138
Db	6696	QIQSDYWRMSLADAPVLLDPTDRSRPSQSFAGALLPVNDLAEFTQSLKRLSEQQVTL	6755
Qy	139	-----	138
Db	6756	FMTLLAAWATVLSRSGQDLVIGTSPSAGRSROEVESLIGFFVNTLALRMDLSDSPVAE	6815
Qy	139	-----	140
Db	6816	LLARVQTALAAQEHQDLPEQVVEIVQPPRLAHTPLFQVMFAWQNNEGMEWMPGLTV	6875
Qy	141	-----	140
Db	6876	SPAQRIDIAKFDLELSLSEVDGRIVGYNLYATALPDQLTIERQVGYLHTVLREMAANPQ	6935
Qy	141	-----	151
Db	6936	QPVGKIDILTVAERKLLLETWNAOTRYPDQTCIHLRFQOBEKTPDAVALVYEHILSY	6995
Qy	152	-----	151
Db	6996	TELNABANRLARQLIKWGISDERVAVILLERSIELIVAOAILKAGAVVPIDPRVDER	7055
Qy	152	-----	159
Db	7056	KNWLINDCSAKLLSGIPVDVAIPRFSLTDEMITIEEDHRDLDRSSTDLAYIMYTSG	7115
Qy	160	-----	159
Db	7116	STGTPKGMVPHRAVRLVINNGYAEIGPDRVVFPEANPAFPASTPEVWAPLLNGTLLV	7175
Qy	160	-----	159
Db	7176	IDHTLLTPKEFVQALQTYRITVLMSVGLFNRLATALFPVLPOIKTLIVGQDVLDPHYM	7235
Qy	160	-----	159
Db	7236	AQVLRDSPQQLNGYGPSEGTFTTTRYITALSPEVSRIPIGRPIANTRVYLLDTYQOP	7295
Qy	160	-----	159
Db	7296	VPQGVTEIYIGDGVACGYLNRPELTAEFLPDPFSDKPDARLYRTGDLARYLPDGNLE	7355
Qy	160	-----	162
Db	7356	FLGRNDQVKIRGFRIELGEIEARLLEYPFVSEAVVQVMGQDKRLIAYVTEADEELV	7415
Qy	163	-----	167
Db	7416	NRLHHLSTILPDYMYPTAFVRLDTPLTPNGKLDRRALPAPGEAFARQVYEAPOGETE	7475
Qy	168	-----	167
Db	7476	ILLAAIWSELLGTEIISRHDSFFALGGHSLAVRMIERLNRNVLGLTLAVRDLFQSPMLSAP	7535
Qy	168	-----	167
Db	7536	AQTVGOSPPVVVPANVITPATTALTPEMLPLIELTOLEIDHIVGQVPGGMANIQDIYALS	7595
Qy	168	-----	167
Db	7596	PLQDGLTFHLLANEGDPYLLVSQMTFDDRALLDRLAAVQVQVNRHDLRTAFVWQGLS	7655
Qy	168	-----	167
Db	7656	TPAQVVRQAPLSVTKLTDPDVGPVRDQALQFDPRRYRIDLQRAPLLHFVVAQBEDGR	7715
Qy	168	-----	178
Db	7716	WSVLQALLHLIGDHTTMEVHMGEVOAYLVQBEENLPVPVPPFRNLVAQVRLGVSAEHTRF	7775
Qy	179	-----	179
Db	7776	FTDMLAEVDEPTLPGLTEVHRDGSQVTSHERMLPVRLNDRLRNQARRGLVSLATLCHLA	7835
Qy	180	-----	179
Db	7836	WAQVLSRTSGQDKVVGTVLFCRMQAGQGDSSMGLPINTLPLRLDMDDTPVQESVRAAH	7895
Qy	180	-----	179
Db	7896	SRLAGLLEHEHASLALAQRCSCIGPEAPLFSALLNRYHSAQPVTTDATISGIEFLGQER	7955
Qy	180	-----	182
Db	7956	TNYPFVLSVEDVGHGLTAQIAQPIDPERICDYMQALESALVALEQHEIPVRTLNLIL	8015
Qy	183	-----	186
Db	8016	PTAEKLLLETWNAOTRYPDQTCIHLRFQOBEKTPDAVALVYEHILSY	8075
Qy	187	-----	186
Db	8076	LAHQIALGVIPOQRVAVCVSRSLAMVGVLAVLKAGAVPLDPVYTGERLTHILTDAA	8135
Qy	187	-----	186
Db	8136	PAILLADNVGRDVLGEDALAGLTVLDPNSPPQPOSNPQVPAALTAQHLAYVYTSSTGT	8195
Qy	187	-----	186
Db	8196	PKGMVVEHRIHLFEATEFTWTFNRQDIWCLFHSIAPDFSVMELGALRYGAKVLVPH	8255
Qy	187	-----	186
Db	8256	AIARSPQELHQFCQHGITVLNQTPSAKFAFIASYIANPLDCLRYIIFGGEALDPSILK	8315
Qy	187	-----	186
Db	8316	PWYALREETLPQLVNMYGITETTVHYTRALARHDVEQITSPITRIPDLTLYLLDKYSQ	8375
Qy	187	-----	191
Db	8376	PVPLGAVGELYIGGAGVARGYNRPBELTAERFLPDPFSSEPDARMYRTGDLARYLPDGNL	8435
Qy	192	-----	191
Db	8436	AFLGRNDQVKIRGFRIEPEIEARLTEHPAVHEAVVQVMGQDKRLVAVVVAEVDKEL	8495
Qy	192	-----	191
Db	8496	INLRTHLGAILLPDYMPAAFCMDAFPLTPNGKLDRRALPVGENAFARQVYAAPQGET	8555
Qy	192	-----	191
Db	8556	ETLLATIWSSELLGIBQVSRHDNFPFALGGHSLAVRMIERLNRNVLGLTLAVRDLFQSPVLSA	8615
Qy	192	-----	195
Db	8616	FAQTAGQSLEVVVPTNVTPTTALTPEMLPLADLTQSEIDHIIAQVSGGVANIQDIYGL	8675
Qy	196	-----	195
Db	8676	SPLQDGLTFHLLANEGDPYLLVSQMTFDDRALLDRLAAVQVQVNRHDLRTAFVWQGL	8735
Qy	196	-----	195
Db	8736	SAPVQVVRQAPLSVTKLTDPDVGPVRDQALQFDPRRYRIDLQRAPLLHFVVAQBEDDG	8795

QY 196 ----- 195  
Db 8796 RNSALQHQHLLIGDHTTMEVMQGEIQAYLAGQENLPVPVPRNLVAQARLGVQAAHTR 8855  
QY 196 -----NVDSTIYF----- 204  
Db 8856 FFTDMLSEVDEPTLPFGLMEVHRDGSQVMESHOMLTPELNDRLSQARRMGVSLATLCHL 8915  
QY 205 ----- 204  
Db 8916 ANAQVLSRTSGQKVVGTVLPGMQAGQAGDNGMGLFTNTLPLRLDMDTTPVDGVRLA 8975  
QY 205 ----- 204  
Db 8976 HTRIAELLEHEHASLALQRCSEVPRETPLFSALLNVRHSAQPVTTDVTISGIEFLSGQE 9035  
QY 205 -----IGLT----- 208  
Db 9036 RTNYPLVSVEDFGHALGUTVQIVQIDPKRVCGYMQALESAAALBQOPETPVRLLNI 9095  
QY 209 -----AL-----LRYAQRN----- 218  
Db 9096 LPEAEKXLLLATWNATQAPYPDQSCIHRLFEQQAENTPNAIALVYGHILNVAELNARAN 9155  
QY 219 -----C 219  
Db 9156 QLAHWLIGQGVKPDNRVAVILLERSVELVAQALAKAGAVYVPIDPRVDERKYWLISDC 9215  
QY 220 T----- 220  
Db 9216 SAKLLTDTPIDLAI PRCLADEMGAIROEDNRNPDLPSSTELAYIMYTSGTGTPKGV 9275  
QY 221 ----- 220  
Db 9276 MYPHRAVRLVINNGYAEIGPDRAFEANPAFDASTFEVWSPLNGVLVVIDHATILM 9335  
QY 221 ----- 220  
Db 9336 PKELVQALQTHRITVLWLSVGLFNRLAAELSPAPQKILIVGGDVLDPHVIRQLRDNP 9395  
QY 221 ----- 220  
Db 9396 POQLLNGYSPSEGTFTTTTTRITALSPENVVRPIGRPIANTRVYLLDTYGHVPVQGAIGE 9455  
QY 221 ----- 220  
Db 9456 IYIGDGVVGYLNRPELTKEFLPDPFSDPENARLYRTGDLARYLPDGNLEFLGRNDQ 9515  
QY 221 ----- 220  
Db 9516 VKIRGFRIBPGETEAQLKHPAVSEALVLTGNGQDKRLVAVYVABADEMLVNLHRLTHLS 9575  
QY 221 ----- 220  
Db 9576 MILPDYMWPAAFVRLDTFTPNKGKLDRLALPVGGEEDFARQIYAAPLGEMETLLATIMC 9635  
QY 221 -----HSFVLVNMGR----- 231  
Db 9636 ELLGVERISRYDNFFVLGCHSLLAVRMNRIRIAGLIEWPLSTLTFPPVLTTFABAISAQF 9695  
QY 232 ----- 231  
Db 9696 GESDRILPAIVPLSREGQLPLSPAQQLWFLTOFGVSETYHIPLALRLHGRLDIAAQO 9755  
QY 232 ----- 231  
Db 9756 TLDTLFARHEALRSFVAVDGPQIPELLPAASGLMRKYDLRMDPVDVDAQLASLSAQEIE 9815  
QY 232 ----- 231  
Db 9816 TLFDLARGPLIRSSLIQLADDDHVFLLTLHHIVDFGWSVLMRELSSLYTAFLAGQDP 9875  
QY 232 ----- 231

Db 9876 LPPLTIQYDYAAQORQWLASERIOSDYMRARLWADPVLDDLPTDRPRPLQSPAGAL 9935  
QY 232 ----- 231  
Db 9936 LPVSLDAELTQSLKRLSEQQGVTLFWTLAAWATVLSRLSQGEDLVIGTSPAGRGRQEV 9995  
QY 232 ----- 231  
Db 9996 SLIGFPVNTLALRMDLSDAPNVTETELLARVRQTALAAQEHQDLPFEQVVEIMQLPRQLAHT 10055  
QY 232 -----NLPRV----- 236  
Db 10056 PLFQVWFAMQNNENTEWRLPGLEVPVEQVFDVAKFDLELNLFEBEGQIVGYLNYATALF 10115  
QY 237 ----- 236  
Db 10116 DOATTIERQVCYLHTVLRGMVANPQOPVGKIDILITVAERKLLLETWNATEMPPYAHCCIIHR 10175  
QY 237 ----- 236  
Db 10176 LFEQQAESPDATAALMYAGQVLSYAEELNARANRLAHQLIMLGVAPEQRVAICVASSPARI 10235  
QY 237 ----- 236  
Db 10236 VGLLAVLKAGAYVPLDPAYPGERLVHILTDAPVAILADNTGRTALGKVLAAALTVDLP 10295  
QY 237 ----- 236  
Db 10296 NSLPQPDSPNOVSALTPRHLAYVIYTSGSTGIPKGVVVEHRLVNLIOEKIVQPEIHFG 10355  
QY 237 -----PKYI----- 240  
Db 10356 SRMLQFASFGDAGVWETWMLCSGATLAI PADTVRQEPRLVHLYLEQAI THACLTPAL 10415  
QY 241 ----- 240  
Db 10416 LREGTDLPEMTIRPTLIILGEAPSATLLOALSRRATVFNAYGPTETTVCATTWRCPSDYT 10475  
QY 241 ----- 240  
Db 10476 EGVIAIGRTVNTQVYLLNTDQOPVPLGAVGELYIGGIVARGYLNRPDLTAERFLADPF 10535  
QY 241 ----- 240  
Db 10536 SDEPDARLYRTGDLARYLPDGNLEFLGRNDQQVKIRGFRIEPEIEARLIEHLAVSDALV 10595  
QY 241 ----- 240  
Db 10596 LTLGQGDKRLVAVYVVAEADDQLVNNLYAHLRAILPDYMWPAAFIRLDTFTPTNGKLD 10655  
QY 241 ----- 240  
Db 10656 RALPAPGVEAFARQVYAAPQGEIETRLAAVWRELLGIEQVSRHDSFFALGGHSLAVRW 10715  
QY 241 ----- 240  
Db 10716 ERLRAGLTLAVRDLFQSPVLSEFAQTQSPVAVPTNVITSATTKLTPEMLPLIDLQ 10775  
QY 241 ----- 240  
Db 10776 TEIDHIEQVPGGIANIQDIYALSPLODGI LPHHLLASEGDPYLLVGQWVFADRALLD 10835  
QY 241 ----- 240  
Db 10836 LVAVQOVINRHDILRTAFIMQGLSTPVQVWCOALLSITELTLNPIDGFPVDOLAQRFDP 10895  
QY 241 ----- 240  
Db 10896 RRYRIDLKEAPLLHFVVAQEDGRWSVLQLLHLLIGDHTTLEVMNGEVOAYLAGQENLS 10955  
QY 241 -----NGT----- 243  
||:



Db 10956 APVPRNLVAQARLGSGQAETHRTFTDMLAEVDEPTLPFGLTVEHNGSQMAELHRMLTP 11015  
 QY 244 KLNKTRKLRK----- 255  
 Db 11016 ELNRLRSQARRLGVSILAALCHLAWAQLVLSRTSQAKVVGTVLFGRMQVGEVDGSMGL 11075  
 QY 256 ----- 255  
 Db 11076 FINTPLRLDIDETSQDVSRLAHLRLAGLLEHSHASLALAQRCQVQNEAPLPSALLNY 11135  
 QY 256 ----- 255  
 Db 11136 RHNVPALATDAITSGIEFLSGQERTNYPLVLSVEDEGDLGLTVQIVQIDPBERICGYMQ 11195  
 QY 256 -----QAPVKQ-----FEKKAKKT 270  
 Db 11196 QALESAAALEQOPEAPVRLNLPASEKKLLLESNNATQAPYPDQCCHRLPEQAENT 11255  
 QY 271 ----- 270  
 Db 11256 PDAIALIYGEHIFNYAELNARANRLARLIRQIQIPDERIAVLLERSIELVWAQAILKA 11315  
 QY 271 ----- 270  
 Db 11316 GAVYVPIDPSVPDERKHLIRDCSAKLLTDPIDLAIPRFCLADEWGAIRGEDYLNLDL 11375  
 QY 271 --QSTTTTFYSYTTSA-- 285  
 Db 11376 SRSSTELAYIMTSGTGMPKGVMPHRAVRLVINNGVAEIGPDORVAPEANVPDAST 11435  
 QY 286 -----I 286  
 Db 11436 FEWAPLNGALVVIDHATVLPKPSAQVLRNYRITVWLWSVGLFNRLAAELSPVFPQL 11495  
 QY 287 NV-----TTNVTYSITTA----- 299  
 Db 11496 NILIVGDVLDPHVMAQVLRDPSPPQQLNGYGPSEGTITTYIRIATLPSGVRIPIGRP 11555  
 QY 300 -----ARRV 303  
 Db 11556 IANTRYVLLDANGQVPLGVEGEIYIGDGVACGYLNRPELTAEFLPDPPFSDKPDARLY 11615  
 QY 304 STSTIA-YRPDS- 315  
 Db 11616 RTGDLARYLPDGNLEFLGRNDQOVKIRFIEPEGEIARLMBHPAVHEAVLWMSGDQK 11675  
 QY 316 ----- 315  
 Db 11676 RLVAIVVAEADMLVNHLRDHLSAVLPDYVPAAFVCLDAFPLTPNGKLDRRALPAGEE 11735  
 QY 316 -FMKSI----- 320  
 Db 11736 DPARQIYAAPSGEMETTLAAIWRELLSVERISRYDNFFALGGHSLAVRMNRIALGIE 11795  
 QY 321 ----- 320  
 Db 11796 LPLSTLTFPTLMAFAVMSARFDESGEALPAIIPLSREGQLPLSPAQRWFLAQFEGV 11855  
 QY 321 -----MATOLR---DLATW----- 331  
 Db 11856 SETYHIPALRURGRLDIAAWQRALDTLPARHEAURSVPVAVDQGVQVVELLPTEWGLPMR 11915  
 QY 332 ----- 331  
 Db 11916 KYDLRTPDQDAQASFSVQESVDFLARGPLIRSLTVQLADNDHVFLLTOHHIVSDGW 11975  
 QY 332 ----- 331  
 Db 11976 SLGVKSELETLYSAYLNQOPDPLPTIQQIPDYAAWQCWLSAAQIQSQSDYWCCTKLAD 12035  
 QY 332 -----VYTTI----- 336  
 Db 12036 VPVLLDPTDRPRPQOSFAGNTLLISLDAELTQSLKLLSEOGVTLFMTLLAAWATVLS 12095

QY 337 -----RYRON----- 341  
 Db 12096 RLSQEDLVIGTFSAGRSRQVESLIGFFVNTIALRMDLSGEPVAELLARVRQTALAAQ 12155  
 QY 342 ---PF-----CBPSRN----- 349  
 Db 12156 EHQDLPFEQVQVIVQPSRRRLAHTPLFQVMPANQNNENTWRPLPGLVVPVEQAFDVVKPD 12215  
 QY 350 -----RTAVSEPMKN----- 359  
 Db 12216 LEINLFEVGRITGYLNVATALPDQVTIERQVGYLHTLREMVANPQQTGVKIDILTVAE 12275  
 QY 360 -----THVL----- 363  
 Db 12336 LIALGVIPDQRVAVCVSRSPAMVGVLAVLKAGGAYVPLDPVYTGERTLTHILTDAAPAIL 12395  
 QY 364 ----- 363  
 Db 12396 LADNVRDVLGEDALAGLTVLDPNSLPDQPDSPQVPGTLAQHLAVVIYTSGSTGTPKGV 12455  
 QY 364 -----YTIYG----- 374  
 Db 12456 MVEHRHLRLFDATESWYRFNRQDIWCLFHSIAFDPSVMELGALRYGAKLVLPHPAIAR 12515  
 QY 364 -----IRNETP----- 374  
 Db 12516 SPQELHQFVCHGVTLNQTSPSAFKAFIASYVANPLPDCRLRYIIFGGEALSPMLKPWA 12575  
 QY 375 ----- 374  
 Db 12576 LREETSQPLVNNYGITETTVHTVYVALARHVEQTTSPIGTRLPDLTLVLLQYQGPVPS 12635  
 QY 375 ----- 374  
 Db 12636 GAVGELYGGAGVARGYLNRPBELTAERFLTDPFSHIPDARMYRTGDLARYLPDGNLEFLG 12695  
 QY 375 ----- 374  
 Db 12696 RNDQVKIRGFRIEPEGEIARLMEYPAVREAVVLADDDQDKRLVAVYVAEVEELINRL 12755  
 QY 375 ----- 374  
 Db 12756 RTHLGAILLPDYMPAAFVRLTFPLTPNGKLDRRALPAGEDAPARQVVAAPQGETETLL 12815  
 QY 375 -----TDMSSLY----- 382  
 Db 12816 ATVWSELLGIEQISRHDSFFALGGHSLAVRMIRLNRVGLTLAVRDLFQSPVLSAFAQT 12875  
 QY 383 ----- 382  
 Db 12876 VQSPAVVVPANVITPVTTLTPDMLPLDLPQBEIDHIVEQVPGGMANIDIYALSPLO 12935  
 QY 383 -----YNETMFE----- 390  
 Db 12936 DQILFHHLLANKGDPYLLVSQTAFFVDRSLDSYLAVVQVQVNRHDLIRTAFTWQGLSTPV 12995  
 QY 391 ----- 390  
 Db 12996 QVWNRQAPLSTVETLTDPIDGVPVYEQLIQRPDPVRHRLDSQAPLLRFMVVAQETDGRWTA 13055  
 QY 391 ----- 390  
 Db 13056 LOLOHLLIGDHTLEWMNRVQAYLTGREENLSAPVFFNLVAQVRLGVSOAETHRTFTD 13115  
 QY 391 ----- 390  
 Db 13116 MIAEVEDEPTLPFGLTEVHRDGSQVTELHRMLPTLIINLRQARRLGVLSAALCHLANAQ 13175



Db 15336 LPRLDMDTPVQESVQAATHRLAGLLEHEHASLALAQRCGVQNGIPLFSALLNVRHNA 15395  
QY 427 ----LDEIRN----- 432  
Db 15396 PVIIVTDEVTNGIBFGQEQERTNYPFVLSVEDPSNALGLTTQIVQIPDPERICDYMQQALE 15455  
QY 433 ---- 432  
Db 15456 SLAALEQKSETPIRLNLILPEAEKLLLESNNATQALYDQCCMRLFEQQAERTPDAL 15515  
QY 433 ---- 432  
Db 15516 ALMYGEYILYAEINARANRLARRLIGLIGIRSDRRVAVLLERSVELVIAQLAILKAGAY 15575  
QY 433 --FSLRSP-----YVN----- 442  
Db 15576 VPLDLRIPDLRHLWLSDCSAKLLTIDIPVDLIDPLCYFADEVDDIEEBGYLNLDPRSS 15635  
QY 443 ---- 442  
Db 15636 TDLAYIMYTSSTGTPKGMVPHRAVRLVINNGYAEIGPDDRVAFEPANPAFDASTFEVM 15695  
QY 443 ----LTPPE-----HRAVNLST----- 456  
Db 15696 APLNGGTLVVIDHTTTLTPQEFVQTLQAVRVTVMLSVGLFNRLAELSTVLPLQKILI 15755  
QY 457 ---- 456  
Db 15756 VGGVDLPDPAIAQVLRDGPQQLNGVGPSEGTTFTTYRITALSPGTMQIPIGRPIANT 15815  
QY 457 ---- 456  
Db 15816 RYLLDIYGQVPQVGTGEIYIGGCVAGCYLNRPELTAEFLVDPFSDLPARMYRTGD 15875  
QY 457 ---- 456  
Db 15876 LARYLPDGNLEFGRNDQKIKRIFGEIEARLVEYPTVQEAIVLVLDGQDKRLVA 15935  
QY 457 ----SNSL----- 460  
Db 15936 YVVAQDEGLANSHTYLSAILPDYMWVPSVFRVDDPPLTPNGKLDRLHALPAGNEAPAR 15995  
QY 461 ---- 460  
Db 15996 QVYEAPOGEMETVLVAIWCELLAEIQVRYDNFFALGSHSLAMRMINLAANHGLICTLN 16055  
QY 461 ---- 460  
Db 16056 DLFQFPVLSLAAKMTSDKLSQPRNSAISVRSQGTGLPLFFVPSGMGDYSYVFGLSQQLH 16115  
QY 461 ----W----- 461  
Db 16116 LGYPIYALPWSISEBPMTEEQATEMINFMKAVQEPGYRICGYSSGGILAYATAQQL 16175  
QY 462 ---- 461  
Db 16176 LYSSEVNVFLGLIDTPAPHYRQRTTPKLPQFIELARQAEQGYIEEMAALYRRIDELNL 16235  
QY 462 ----W----- 462  
Db 16236 VQFIETAQKALYANLADLIAKRWEQIANYAQIVRDYEPQVSAITLHQFYAMESSPSV 16295  
QY 463 ----WLG 465  
Db 16296 PVVIDTKIMDIEPQPLNMGSSLGWAQ 16321

RESULT 7  
ID ABB84277  
XX ABB84277 standard; protein; 11300 AA.  
XX ABB84277;  
XX

DT 08-NOV-2002 (first entry)  
XX Adenovirus Adilp associated protein SEQ ID 2.  
DE  
XX  
KW Genome; cancer therapy; vascular disease; cytostatic; vulnery; cancer;  
KW liver; lung; breast; prostate; bladder; glioblastoma; neuroblastoma;  
KW medulloblastoma; hepatoma; adenoviral vector; gene therapy.  
XX  
OS Mastadenovirus.  
XX  
PN WO200253759-A1.  
XX  
PD 11-JUL-2002.  
XX  
PF 04-JAN-2002; 2002WO-SE000013.  
XX  
PR 04-JAN-2001; 2001SE-00000035.  
PR 08-JAN-2001; 2001US-0260358P.  
XX  
PA (WADE/) WADELL G.  
PA (MEY/) MEI Y.  
PA (SEGE/) SEGERMAN A.  
PA (SKOG/) SKOG J.  
XX (LIND/) LINDMAN K.  
PI Wadell G, Mei Y, Segerman A, Skog J, Lindman K;  
DR WPI; 2002-590637/63.  
XX N-PSDB; ABQ76121.  
PT Novel DNA sequence encoding adenoviral proteins, preferably fiber protein  
XX and the hexon of serotype 11 adenovirus for gene therapy of vascular  
PS diseases and cancer therapy.  
XX Claim 1; Page 52-97; 102pp; English.  
CC This invention describes a novel purified and isolated deoxyribonucleic  
CC acid sequence representing the Adenovirus type 11p genome. This genome  
CC can be used to construct a vector for gene therapy, cancer therapy or  
CC therapy of vascular diseases. The product of the invention also has  
CC cytostatic and vulnery activity and can be useful for gene therapy of a  
CC patient suffering from cancer such as liver, lung, breast, prostate and  
CC bladder cancer, for the transfection of human cells such as cells of  
CC neural origin, glioblastoma, neuroblastoma, medulloblastoma, hepatoma,  
CC breast cancer, endothelial cells and dendritic cells, and for infecting  
CC human endothelial cells or hematopoietic progenitor cells. The invention  
CC is also useful in vascular surgery, for the formation of neo-intima in  
CC vascular prosthesis, such as vascular stents and in the prevention of  
CC scar formation in surgical applications. The adenovirus vector described  
CC in the invention exhibits a high tropism towards endothelial cells and  
CC prevents the formation of neo-intima and microangiogenesis. This sequence  
CC represents a preliminary amino acid sequence associated with the  
CC Adenoviral type 11p Adilp genome described in the disclosure of the  
XX invention  
SQ Sequence 11300 AA;

Query Match: 23.7%; Score 580; DB 5; Length 11300;  
Best Local Similarity 2.2%; Pred. No. 0.00014;  
Matches 233; Conservative 94; Mismatches 137; Indels 10068; Gaps 73;  
QY 2 GRKWM----- 7  
Db 84 GCKKLLIPARKLNEEVFFIMWYWGQVFGVQVDFDPLRGGDFYRVFYLNFRVPCQSL 143  
QY 8 ----- 7  
Db 144 FLRRCOLIARVFIPOGLCOEATLEQREBFSPLRKQFNKNKMRDLRFLPQEIISAETGNE 203  
QY 8 ----- 7  
Db 204 ILELVVHALMGDDPPFPVQLFEPPTLQELYDLEVEGSEDSNEEAVNGVFTDSMLAANEG 263

QY 8 -----VRDVPKM----- 14  
Db 264 LE LDPPLDTPTDTRGVIVSGTGVKLPDLSSVCDLHCYEDGPPSDEEDHEKEQSMQTA 323  
QY 15 -----FV----- 16  
Db 324 AGEVKAANVGQFQDCELPFGHCKCEFRKNTGVKELLCSLCYMRTHCFYISKCVVK 383  
QY 17 -----LISISFLL----- 24  
Db 384 IRNMLFETCILSVFVLLIIGPVSDADESPSPSTTSPPEIQAPVPDVVRKPIPVKCLKPG 443  
QY 25 ----- 24  
Db 444 KRPAVEKLEDLQGGDPLDSTRKPRQVFIHVRVLRQYLCD SAMKYVNCSLVFI AFW 503  
QY 25 ----- 24  
Db 504 AGTQVYKQTCVVSSSELAFTHGGLGHGRPEDATVRFGRSLRFLLEILVRISGSPDKTG 563  
QY 25 ----- 24  
Db 564 LTRIKVUGRLPRTFSSFGSGSLRKSFI SFRLPNPRNCCCPSFYIRMDPADSFOQGI 623  
QY 25 ----- 24  
Db 624 RFGHSHSIVENMEGQDEDNRLASAAFGCSGNPEASTGHASGGGTARGQPESRPG 683  
QY 25 ----- 24  
Db 684 PSSGGGVADLSPLOVLVTGSTGRDGRVKRERASSGTDARSELALMSRRRPETIW 743  
QY 25 ----- 24  
Db 744 WHEVQKEGRDVSVLQEKYSLEQVKTWLEPEDDWEVAKNYAKIALRDPKQYKITRRIN 803  
QY 25 -----VSPINCKV----- 32  
Db 804 IRNACYISGNABEAWIDTQKAVIRCCMDMWPVGVGMEAVTFNVKFRGDGNGIVFMA 863  
QY 33 -----MSKALYNR----- 40  
Db 864 NTKLILHGCSFFGFNNTCVDAMQGVVRGCSFYACWIATAGRTKSLSLKCKCIFORCNLG 923  
QY 41 ----- 40  
Db 924 ILNEGEARVHCASTDTGCFILIKGNASVKHNMICGASDERPYQMLTCAGHCNMLATVH 983  
QY 41 ----- 40  
Db 984 IVSHQRKKWPVFDHNVMTKCTMHAGRRGMFMPYQCNMNHVKVLLBPDAPSRMSLTGIFD 1043  
QY 41 ----- 40  
Db 1044 MNQIWKILRYDDTRSRVRACEGCGKHARFPQVCVDVTEDLRDHLVIARTGAEGSGGE 1103  
QY 41 ----- 40  
Db 1104 ETDGEYENFGVFSQDQIEKFPVFSVLQSVETILLRGESSALIQVSHPGQEFVRMLWD 1163  
QY 41 ----- 40  
Db 1164 LLWMEPSNPPILQRPMLLVLHLWTQLPPLPLPLTLCLWVVTMEASWLIPLPLITLL 1223  
QY 41 ----- 40  
Db 1224 PLRTSYLFWPSWRLPNVWNFLRWSSEYKLSLSARQSLNKKIPESMNKTSLLLINQ 1283  
QY 41 ----- 40  
Db 1284 VFLFHSRTVCPRPPIISIENSVDFFQDPIEVGLNVHGHVAVGVEIAPLKGFMRLGSVV 1343  
QY 41 ----- 40

Db 1344 NHPVITRSCQVMLHNI FKADCHRALGVGVYKPVELGWVHSGVNYVHGLDFVGNIAAKIPS 1403  
QY 41 ----- 40  
Db 1404 WYHVMKDHDGVSGTFRKPIVOLGWKSVKEKFGDTLVSSKIFHALIHDSNGAVSGAGKH 1463  
QY 41 -PW----- 42  
Db 1464 VZWVHIIVMFLSIIISHFNEFGAESTRLGYECSPGRSIVPLTDLHFPSPQPRGMNHVHL 1523  
QY 43 ----- 42  
Db 1524 GGYEKHFWGGDLQISEOLREATSGGAINDSYGLQVVVGTTATAVFSKQGGHLVHHFPY 1583  
QY 43 ----- 42  
Db 1584 MHI FPHQIHEALSSKPLGKVFORFOTVSHGHGESSLLQKFSVPQFSDVPYGISIQOTSSF 1643  
QY 43 ----- 42  
Db 1644 RQFGRLLBGMREWASAAARVRSFQGLSVRVVVTVKGCAPAWALARVPRLLILLVNF 1703  
QY 43 ----- 42  
Db 1704 CRLAPCMSAKOFTMSSLSASAAWPLARSLPLEVFLHTQYRHFSAYNLGARKTDSGEVAS 1763  
QY 43 ----- 42  
Db 1764 APOEAQTVSHSTSVQVKSGLSGKTSFPFPYFLMRFLPLVMSMSPRVNRLSVSPTDFTGL 1823  
QY 43 -----RGLVLSKI----- 50  
Db 1824 PSSGVPRSSYRNSDHTKARVQASTKEAMWEGRSLSRGSTFFSKVCKHMSPSSTRNV 1883  
QY 51 -----GKYKLDQ----- 57  
Db 1884 IGLVYFTQVPAGVKGAVLCSSLSGSLSRNVSCWGRYLSKAGMTSALRLSVSKNEE 1943  
QY 58 ----- 57  
Db 1944 DLILTVFEMPMPRSSIWSENTIIFLLSSLVANDPYRALDKSLAMDRMVWFFSLSARSLA 2003  
QY 58 ----- 57  
Db 2004 AMLSMTYSRARHFHSGKIWNSSGTLTCHPRLCKVIKSTLVATSPRRGSLVQOQLPPL 2063  
QY 58 ----- 57  
Db 2064 EQKGGSGSISSSGGSGASMWKIPGSKSLSKLMGVGSKAICHRAASARSYGLRGLPHGM 2123  
QY 58 ----- 57  
Db 2124 GWVSABATPQMTMGSSKMPMVGHRPPLILARTSYSCDGSASPQKLVRLGFSVLTIW 2183  
QY 58 ----- 57  
Db 2184 RQKAELEEMVGLKMLKWAGRPTESLT KWADSSLVTSSAVTSTSRAQSSVSMMSPGWFFPS 2243  
QY 58 ----- 57  
Db 2244 HSSRLRYSRSRFOYSSSGNPSLSARDPSMNL TALGQOPFSTGREYAAAFRSEAVRAKVS 2303  
QY 58 -----LKLILRQ----- 65  
Db 2304 LWTULRNWYLSMSQAPCSQSWKSTRFLAGLKAKVTSLKRLPALGIKLVWRKCGCT 2363  
QY 66 -----LETITISK-----YN----- 75  
Db 2364 SARLLITWAARTISSKPLMLCPTMYNMRGVPLTGLSSKVSRSVSGDKACSAHSCR 2423  
QY 76 ----- 75

```
Db 2424 GFACRNDQBSTASAVCNMSRYKCRPIAIFSGVTOKVLGSCCHSHLSILMARSWMLTR 2483
Qy 76 -----VSKOPV-----KN----- 83
Db 2484 RSSPESFMTSMKTSCLPKDPIQVWSTSVKSLSVRGEPIGKMWISCHQLEDWLLMKKF 2543
Qy 84 -----LTMN----- 87
Db 2544 LRAEHSCLLYRRPOSQRTGCSMSCTWLPLTRNFGSKPRGDCISCSSIFAVSACS 2603
Qy 88 ----- 87
Db 2604 SVSMVWLTSRGRQVQTSAREGRSTRARRLELSVLRCGLRVGRDRRLTCMIFSR 2663
Qy 88 -----TEFPQYILAG----- 98
Db 2664 CGRFRWLSTGTSFVETSMACRVPCLGATTVPLFLLIGGSLASCMRLSGDGDARRAA 2723
Qy 99 ----- 98
Db 2724 AVVPDPRAMLVWARRRRRAGSGTALEDLRAAPPVDRVLDVSGKLPAPATKRVOQOQPR 2783
Qy 99 ----- 98
Db 2784 YRQLVSVFLRHQSCFGRSPPTARFLPDLRDLFRWRPGRHWYGPVGRMHSCPPR 2843
Qy 99 ----- 98
Db 2844 SRRGCKPRPSRLRASPPERGAPVRWPHSCIGAEGKSVWQCVRRRNTSIVSAAFRH 2903
Qy 99 -----PIQH----- 102
Db 2904 RPELSGAPRRSPRONKTGSPARTSIPREDGVLWMPVLVRLPGSLPLLSLLEPL 2963
Qy 103 ----- 102
Db 2964 TSLRLQAGAEAGDVGARANGRIVQPLRGGAWFQRRGRSRAVAEKHRRASPSGDWE 3023
Qy 103 ----- 102
Db 3024 VLRLGGRGRLYILLIGPLHAEISCQDPRDLKTRFRKRLTSHHKVGVRLLVGGGYVFG 3083
Qy 103 ----- 102
Db 3084 LGLLLHLLGKVRCCWNSRQFDGWGAPGLWVRLAGYAGWPFPKHPYDQIDLCSSL 3143
Qy 103 ----- 102
Db 3144 AAVLRALLPHFPCHAYVQIRALVVPVPSQLRFLRRGWLAVLGGWLESHQNPQSGGKLLY 3203
Qy 103 ----- 102
Db 3204 WKGHSWPLTSLSDQAGARCIGANRRGRCRCRCPDTGTLENAAVVGGREAILVLE 3263
Qy 103 ----- 102
Db 3264 RQGRGLPTGDSRRCTWTSRFLARKPEETRVGSKCVAASSLARFDQAGRSHCSIDTE 3323
Qy 103 ----- 102
Db 3324 KMKAFSdstppGGTTGWAVYPGSRLVLEPAGAAANVVLALPSRPSLQKSRIRNRVLLV 3383
Qy 103 ----- 102
Db 3384 SEWQGESYFFPPAAQHPVLQRMPQQPPSQOQQOQSQKAVPATTATAAVSGADSPM 3443
Qy 103 ----- 102
Db 3444 IWTWRAKDWVVRHPGSGIREFNKKILARRMCPNRTYLEAARRRRRCELPAITRVVS 3503
Qy 103 ----- 102
Db 3504 CVTVWTEDECCGTRISKLMKQSVLPFGHTWLOPTLYRLTSRQRKS VTSKSLLIIMCEPLP 3563
```

---

```
Qy 103 ----- 102
Db 3564 AKKLPLVICIGIWKLSFRTLNLPPSCFWCWNATMELSERCTSPNPRGDCMILST 3623
Qy 103 ----- 102
Db 3624 FYVSCRSRAWAPRRWLPISITRPAWENITLAKSTRLHTPTPTRRMGSTCARSRSPAMIL 3683
Qy 103 ----- 102
Db 3684 GCIAMTECIARLAPAGASATGCTVCKELLEPRVITSTWELTCSGLVAGLAPRRQ 3743
Qy 103 ----- 102
Db 3744 DVSFLTTRMKARRKRASTWKTGTTRVFCMEQAPDPAMRAALQSPGINSDDWTOA 3803
Qy 103 ----- 102
Db 3804 MORIMALTTRNPEAFRQOPQANRLSAIMEAVVPSRSNPHEKVLATVNALVENKAIRPDE 3863
Qy 103 ----- 102
Db 3864 AGLVYNALLERVARNSSNVQTNLDRMITDVREAVSORERFORDANLGSILVALNAFLSTQ 3923
Qy 103 ----- 102
Db 3924 PANVPRGQDYTNFLSALRLMVSEVPQSEYQSGPDYFPQTSRQGLQTVNLSQAFKNLKG 3983
Qy 103 ----- 102
Db 3984 LMGHAPVGERATVTSLLTPNSRLLLLVAPPTDGSIDRNSYGLVLLNLYREAIGOSQV 4043
Qy 103 -----YSIT----- 106
Db 4044 DEQTYEITQVSRALGOEDTGSLEATLNFLLTNRSQKIPQOALTAEERILRYVQOSVG 4103
Qy 107 ----- 106
Db 4104 LFLMQEGATPTAALDWTARNMEPSMYASNRPFINKLLDYLHRAAAMNSDYFTNAILNPHW 4163
Qy 107 -----YLM----- 109
Db 4164 LPPPGFTYETDMPDPNDGFLDDVDSDVFSPLSDHRTWKEGDRMHSSASLSGVMGAT 4223
Qy 110 ----- 109
Db 4224 AAEPESAPFPLPSLHSHSVRSSEVGRISRPSLMGEESEYLNDSLLRPAREKNFPNNGIES 4283
Qy 110 ----- 109
Db 4284 LVDKMSRWKTYAQDHRDEBPGINGITSRASRRRQRHQRGLVWDDSDSADSSVLDLGR 4343
Qy 110 ----- 109
Db 4344 GRGNPPAHLRPLGGLMKIKKLTAMATSVRSFFIICVYNEASRARRSGVSGSSS 4403
Qy 110 ----- 109
Db 4404 FVRERDAAAAGGGDAIPTGGSLSASAPGTGGOKQHSFGTGTSTVRHYHVQVSGQVQV 4463
Qy 110 ----- 109
Db 4464 GHCPSELSEPOLLDHGGAKQLYPYGSQHPDHLTIAGVRSKXDHAYHAKRERVYVQVS 4523
Qy 110 ----- 109
Db 4524 ACDGVQKTSRRCCSWGYSQAGYFBIYVVRYPARQLFSYYDYFDEQCHHRLLESSTWS 4583
Qy 110 ----- 109
Db 4584 AKHWCVRHQELOAGMGRNQVDHANSVYVLSHCLTAWLRSGFYRESFEQSPSYQKTA 4643
```



Db	5804	HGPRARLCHCKYGRDGGRALVGRLEPTFLLPFFSFWILGSSQDLPVIGSPAPORSCY	5863
Qy	220	-----	219
Db	6864	QGPLYYAGKIYPDRAGSPFCRLRTEFLHVPRLCALAPSHGRKPHEIANWSAKHASFS	6923
Qy	220	-----	219
Db	6924	SPAHPVQSKSTLPSPYPFALFSLPSYTHRGHCVCYPYGCIMTHVNVNFKHHFIFLHVS	6983
Qy	220	-----	219
Db	6984	RLCITYLTSRMGSDENQNDPOAIVIRCGTDTWTATIRESPTEVPVYRAGCHSTAFWSAAK	7043
Qy	220	-----	219
Db	7044	LQAGQPKSNHNDQCFERESCGTDCSTETPSATDVSRLLPARWDLQCPHPDLQHWQCTG	7103
Qy	220	-----	219
Db	7104	SSCRSAYPWRAPNACGNRAGGSVSWDPDVFELTRLXKHITAKPAGLYPYRYKTSRRT	7163
Qy	220	-----	219
Db	7164	CSKTGLSRHSHSSGRHCLFAPHFCPSGFGFWFARDSPRLVLVRSRWPHPSRAPSSEY	7223
Qy	220	-----	231
Db	7224	CHAGTSACPHNCHSHEATHTSLYIPYVGRSEKKNVFPABIFPPSSCVSCDKLTGCLGA	7283
Qy	232	-----	231
Db	7284	PRLRTGDCACIVRVAQALVRFVYPACTSPSADTSLPCLSPKQTPGASSDSQCRQOLL	7343
Qy	232	-----	231
Db	7344	PEGHLRSSQCFCHPSQRCARAGSNPLQVAPLLFLRLCLDCLAWGVVWSSLASFWGVSE	7403
Qy	232	-----	N 232
Db	7404	EEDCRSPETGRIVTFPSPLPTDCRKNLTPHGDRCFSFGAEVIAKAGCGPTWKADWQN	7463
Qy	233	LFRV-----	GTKL 245
Db	7464	PFRVRCAPCGGLTDFLRGWPLCSQRNRHNGSAIAVNIATSAITSRPPQRKGAEI	7523
Qy	246	K-----	246
Db	7524	KHSTAOSCHHLYPRRGGRIISHAEKSERVDRHRARPGLCDTGTGRVETLSRERKLPK	7583
Qy	247	-----	246
Db	7584	TTSLRSPRCWKGSEHLPHRARRAPTSSTKTVAHSGGCIIGNSAHQCGRAQPLRAPL	7643
Qy	247	-----	246
Db	7644	FTSYSPQSAKRLRAKSLKLLSPCCARSTGYLSHLFSKNSLLPRSHPRRCPTQSG	7703
Qy	247	-----	246
Db	7704	TWFTLTYFLGRGSKDLRGSGDGRKCSAKGRKWHGASQSGGIGRRQCOTRSTQAKHRG	7763
Qy	247	NTWR-----	250
Db	7764	HTLIRSQCQAPSHDGHGVPVTHARKSPFRHAPRCLGTSGQANPMAGHRLSQGFGRAS	7823
Qy	251	-----	250
Db	7824	QAYDGRGAGYRTRVSPTEFLYPRNLAQTRRESALHFTRLCAAGMODIRGTHQPGFLHY	7883
Qy	251	-----	KLKKAQ-----
Db	7884	SAESPRTKRAAQHEGGSPPLHPRLCLSVFPVNVANRHGMAAMPRTTELEAQALETIS	7943
Qy	258	-----	257
Db	7944	GSVDRVRRARRFRPGRPHLPASQGYFAKRIALYEPHAQFSLFHPGTILYPARHLLRT	8003
Qy	258	-----	257
Db	8004	ALRLCASHLPVPPPAAMESLLVPBQSLPLPGDCRGCCERRRLAGVSLPQSVHAPPVP	8063
Qy	258	-----	262
Db	8064	SLQPPVDERNPNRHLIARPPQPRRWVFSWAKEFTDPGTVDLRLLLAQVCSGRLLPPLNQVL	8123
Qy	263	-----	262
Db	8124	GPITASKRGTRFGLRHHPGGNSGPIASHKPIPPRISTEKGGGLPPPPDRRGTOHKVPSCPN	8183
Qy	263	-----	PEKKAK-----
Db	8184	DEKTRSCRRRPQKIWRKIGTVRQRRRTVWRTVWRKTRQRRRKKPPTNSYP	8243
Qy	270	-----	269
Db	8244	RLRQATALPSLRVEEPCGVPAVDGTRPDASRTQPALPRVRRIGRDTSPGGGIRPSS	8303
Qy	270	-----	TOSTTTPYFS-----
Db	8304	PACMSAGATYPSRGATCYSTMGTFRAMFCITVTSTAPTIAKSRQSRQIKTAAATSNRK	8363
Qy	280	-----	VTYSI-----
Db	8364	PAAVRKYTTSAATGGLKITANEPAQTRELNRIPFTLYAIFQSGRQGBQELKIKNRSUR	8423
Qy	297	-----	296
Db	8424	SLTRSLYHKSQDQORTLEDAELFNKYCALTLKEAATALLQKRRLHHPHRRNSHAL	8483
Qy	297	-----	296
Db	8484	HVELSTPNCIGRRLLPGLLHPHQAORAFYDFSSYTRLPKPTFCGTSSYHAPTPSQ	8543
Qy	297	-----	VTSTIAR-PP-----
Db	8544	KLARRPSVEGKSRSHHCITSSRRPGRSPNDCRCAVSWRLHPMSSQASAYKTDDQPRYP	8603
Qy	314	-----	SSFMKSIATOL-----
Db	8604	AQRRVGLSAMSTTRNLSDCRLREIFLHPSSGCCDFGKVFATPLGRNDRSICGVYS	8663
Qy	330	-----	VTTLR-----
Db	8664	LCLLQPLLRISWALPGRVHTELRRDRVSGRLMSGDAEALSLRLHLDHCRPRCPAREL	8723
Qy	338	-----	YRNP-FCEPSRNTAVS-----
Db	8724	IBFIYFELPKDHPQAGHVRISIEGKIDSLRQIFSORPVLIERDQNTTVSIYICNH	8783
Qy	355	-----	EFMKNTHVLR-----
Db	8784	PGHESLCLMCTEFKNKIKTLRATAASTRILOPERNFSCRPGCLHLHSYQTRSSYT	8843
Qy	368	TP-----	369
Db	8844	TFLQKHPFYFQNRAPRSSYRKPLGGSGPCARNSCGWACDYSLLPIHTLLHLSGV	8903
Qy	370	-----	369
Db	8904	VVLVRMGPIVLVLLLSLLEPGSANYDCLDFDPENCLTTFAPDTSRICGVLIKCGWECR	8963
Qy	370	-----	YTI-----
Db	8964	SVEITHNNKTNNTLSTTWEPGVPEWYTVSVRPGDGSIRISNNTFIFSEMCDLAMPMSQ	9023

QY 373 ----- 372  
Db 9024 YSLWPPSKDNIVTFSIAYCLCACLITALLCVCIHLLVTTTRIKNANKKMPPLSVYRHGF 9083  
QY 373 ----- 372  
Db 9084 SYISHICQHCHCRSWNSRLYPSRTLLSHRTPNFRHLGQTGKRLLYNLQNKTNNSNL 9143  
QY 373 ----- 372  
Db 9144 QHTKSYIDCSLQRLLLWLQIQSIKLLGSCYPVENHENAKYKDSIQFQRNFIYSHHTR 9203  
QY 373 ----- 372  
Db 9204 RKKPRFNDCCSGSGDGTNNMHAFICLSLQKVSSKTRSPTKAHLISFYTAMYSTTT 9263  
QY 373 ----- 372  
Db 9264 FLMLTSLATLSARSHLTVTIGSNCTLKGPGQGHVFWRIYDNGWFTKPCDQGRFFCNG 9323  
QY 373 ----- 396  
Db 9324 RDLTIINVTANDKGFYGTIDYKSSLDYNIIVLPSTPAEPTTTTFSSSVANNTIISPTFA 9383  
QY 397 ----- 401  
Db 9384 ALLKRTWNSSTSHTTISTSTISIIAAVTIGISILVFTITYYACVYKDKHKGDPLLRFD 9443  
QY 402 ----- 401  
Db 9444 IFVLFFIYVMVNTNHTKFLHLHHTLCICLRYFSSSHSNPRLYRSICFLCTFCFCYLH 9503  
QY 402 ----- 406  
Db 9504 LRMHSLPGYFFPYRLDPCANCLPAPSPRIQPKYRGSTHLKPCRLYYQYFCYCFPTL 9563  
QY 407 ----- 406  
Db 9564 SOPQLPIVLHONTLENANSNNRGHFLLAIEKKNQKPPQILLELISVAPPHFYTPYLILAGM 9623  
QY 407 ----- 406  
Db 9624 LPMHMIHKTORWTFYKTCNIQRIKVNHNPHYSLLLVSTPAEMTETLTTSNSAEDLL 9683  
QY 407 ----- 406  
Db 9684 DMDGRVSEQLAQLRIRQQERAAKELRDVQIHQCKKGFICLVKQAKISYEITADHRL 9743  
QY 407 ---MGFOR----- 411  
Db 9744 SYELGPQRQKTCMVGINPIVITQSGDGTKGCHCSDSIECTYTLKTLGLRLDLPWN 9803  
QY 412 -----TF----- 413  
Db 9804 KMINKSLTNQOGLCNFLPAAPHPLPNSGILNPVQRTFSLRGQILAPLLYQSSCL 9863  
QY 414 -----IDPLWDYLD----- 422  
Db 9864 SSQMTKRVRLSDSFNPVYFEDESTSQHPFINPGFISPNGFTQSPNGVLTCLKLPLTTT 9923  
QY 423 ----- 422  
Db 9924 GGSLLQKVGGLVDDTNGFLKENISATTPVLKTGHSIGLPLGAGLGTNENKLCIKLGQG 9983  
QY 423 ----- 422  
Db 9984 LTFNSNICIDNNINTLWTVGNPTEANCOIMSSSESDCKLILTLVKTGALVTAFFVYVG 10043  
QY 423 -----SLFLDEIRNF-----SLRSP----- 438  
Db 10044 VSNFNMLTTHRNINFTAEFLPFDSTGNLLTLRLSLKATPLNKHSGQNMATGAITNAKGFMP 10103  
QY 439 ----- 438

Db 10104 STTAYPFNDNSREKENIYGTCTYATSDRTAFPIDISVMLNRRRAINDETSYCIRITWSWN 10163  
QY 439 -----TYVNLTPPEH----- 448  
Db 10164 TGDAPDEVOTSATTLVTSPTFFYYIREDDQIKFNLFIKSIHKIRVWILPPPSHLTEYTNLS 10223  
QY 449 -BRAVN-----LSTNSLW----- 462  
Db 10224 PRTALNINIPLDIDMWLDSTFQTVSERANLGSVIDKNPSGFKALSQSCCGTPESSGR 10283  
QY 463 ----- 462  
Db 10284 SSGRRTWGIIIRKRYRTIVSHQTHKQPLSASLRATAVYGIHVHSLKHDNFSPHQLSGAM 10343  
QY 463 ----- 462  
Db 10344 RAATHSDFTQIFAVGTTHYINVTIISAPAKTHIYNRPCMTIIPKENINMTPPQKHHTH 10403  
QY 463 ----- 462  
Db 10404 IHDLFWHVHNNLSVPWTTLVNHAQYNLPEPHCQHRSPSHALKTLLITMTWKNPILSTV 10463  
QY 463 ----- 462  
Db 10464 NHLRMKNYSGTTTMMHASSHNFLLRIKIPGNRKLQNSKAGTRKTNTVTVMHSHSIT 10523  
QY 463 ----- 462  
Db 10524 IWQORVFSHRSSGFIFLTTLWGSVVRVMSGACRACAOQCHNGVASHSRILYSKTRPWQN 10583  
QY 463 -----WLQ 465  
Db 10584 TLFFAFYPAARVPCDSSSTTILLSKECWLQ 10615  
RESULT 8  
AAR44929  
ID AAR44929 standard; protein; 15281 AA.  
XX  
AC AAR44929;  
XX  
DT 16-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 08-JUL-1994 (first entry)  
XX  
DE T. niveum Cyclosporin synthetase.  
XX  
KW Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;  
KW T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase.  
XX  
OS Tolypocladium inflatum.  
XX  
PN EP578616-A2.  
PD 12-JAN-1994.  
XX  
PF 05-JUL-1993; 93EP-00810474.  
XX  
PR 09-JUL-1992; 92AT-00001403.  
PR 08-MAR-1993; 93AT-00000437.  
PR 29-APR-1993; 93CH-00001310.  
PR 04-MAY-1993; 93CH-00001375.  
XX  
PA (SANO ) SANDOZ LTD.  
PA (SANO ) SANDOZ PATENT GMBH.  
PA (SANO ) SANDOZ-ERFINDUNGEN VERW GES MBH.  
XX  
PI Leitner E, Schneider E, Schoergendorfer K, Weber G;  
XX  
WPI: 1994-010432/02.  
DR N-PSDB; AAO54386.  
XX



PT Isolated DNA sequence - which codes for enzyme having cyclosporin  
 PS synthetase like activity.  
 XX Claim 1; Page 41-84; 93pp; English.

CC This sequence represents an enzyme which has cyclosporin synthetase- like  
 CC activity. This sequence was isolated from Tolypocladium niveum (formerly  
 CC known as T. inflatum GAMS). This enzyme catalyses the peptide  
 CC biosynthesis of cyclosporins and structurally related molecules. This  
 CC sequence may be used for the production of cyclosporin by transforming a  
 CC vector containing this sequence in to a recombinant host. This allows  
 CC effective production of anti- biotic cyclosporin or its derivatives.  
 CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 16-OCT-2003 to  
 CC standardise OS field)  
 XX SQ Sequence 15281 AA;

Query Match 23.5%; Score 575; DB 2; Length 15281;  
 Best Local Similarity 1.7%; Pred.No. 0.00041;  
 Matches 254; Conservative 78; Mismatches 128; Indels 14673; Gaps 82;

QY 1 MG----- 2  
 DB 1 MGAIGQMAYDLANPSRASSISSNRYSEPVQSPAQRGLWFLHQLKLGASWDITPAAIR 60  
 QY 3 ----- 2  
 DB 61 LRGHLDIDALNAASRALTORHETLRTTFKEQGVGVVHASSGLERGLRIVDASSRDIAQ 120  
 QY 3 ----- 2  
 DB 121 LLABEOTMKFLESEPAWRVALLKVAEDHHILSVVHHIISDSRSLDIIOBELGELYTAA 180  
 QY 3 ----- 2  
 DB 181 SQKSISACPLGPIQYRDLTWQNDQVQAEQERQGLGWIEQDNNTPAELLTELPRP 240  
 QY 3 ----- 8  
 DB 241 AIPSGETGKISQIDGSHKELLAFCKRSQQVTAYAVLLAAFRVHPRLTGAEDATIGAPV 300  
 QY 9 -----RDVP----- 12  
 DB 301 ANRDRELENMVAPLATQCMRVVLDEDDTFESVLRQIMSVMTAEHANRDVFFERIVSAL 360  
 QY 13 -----KMFVL----- 17  
 DB 361 LFGSTDTSRHPLVQLMFALHPAQDTGRARWGFLEAETLQSAAPTRFDMEMHLPFGDDREF 420  
 QY 18 ----- 17  
 DB 421 ANVLPSTGLFDAEAIKRSVVSIFREVLRRGISBPVHVKTMTPLDGLAARDMGLLDIGTT 480  
 QY 18 ----- 17  
 DB 481 DYPREASVDMFOEQVALNPATAVADASSRLSYSELDHKSQDLAAWLRRLQKPELIG 540  
 QY 18 ----- 17  
 DB 541 VLSPPSCETWSPFLGILKAHLAYPLDINVPALARIESILSAVDGHKLVLLGSNVPQKVD 600  
 QY 18 ----- 17  
 DB 601 VPDVELLRISDALNGSVNGLAGKQATAPKPSATDLAYVFTSGTGKPKGVMEHRRGIVR 660  
 QY 18 ----- 17  
 DB 661 LVKGTNIISPAQAAVPTAHLAIAFDLSTWEIYTPILNGTLVCIEHSVTLDSKALEAVF 720  
 QY 18 -----ISISFL----- 23  
 DB 721 TKEGIRVAFLAPALIKOCLADRPAAIFAGLDSLYAIGDRFDRDALHAKSLVKHGVNAYG 780

QY 24 ----- 23  
 DB 781 PTENSVVSTIYSVSEASPFVTCVPVGRAISNGAYVMDQDQLVSPGVMBELVSDGLA 840  
 QY 24 ----- 23  
 DB 841 RGYTDSALDKNRFVVVVQIDGESIRGYRTCDRARSLSKGGQIEFFGMRDQOVKIRGRIEP 900  
 QY 24 ----- 23  
 DB 901 AEVEHALLNSDQVDAAVVIRROBEEEPAMIAFVTTQGTLPDHLVNINGHVPDGNBSK 960  
 QY 24 ----- 23  
 DB 961 NDQFAVHVESELRRRLQMLLPSYMPARIVLDHPLPNPGKVDKALCQSAKTQVQSKL 1020  
 QY 24 ----- 23  
 DB 1021 VSORVAPRNEIEAVLCEEVRSVLGVEVGITDNFPLDGGHSLTAMKLAARISQRLDIAQSV 1080  
 QY 24 ----- 23  
 DB 1081 ATVPEQPLADLAATIQRGSTLYSVIPTTEYTPVEQSPAQRGLWFLQELNTCASHWYVM 1140  
 QY 24 ----- 23  
 DB 1141 LTVRLRGHLVDALGTALLALEKRHETLRTTPEERDGVGMQVVHSSLMGELRLIDISEKS 1200  
 QY 24 ----- 23  
 DB 1201 GTAHEALMKEOSTRFDLTREPGWRVALLKLADHHI PSIVMHHIVSDGSLDLRLHELQ 1260  
 QY 24 ----- 23  
 DB 1261 LYSALRGQDPLSRLEPLIQYRDEFAVWQKDSQKAAHQRLQYWTQKQADSTPAELLT 1320  
 QY 24 ----- 23  
 DB 1321 DFPSPILSGKAGKVPVVAIEGSLYDTLQVFSRTHQVTSFVALLAAFRAAHFRLTGSDNAT 1380  
 QY 24 ----- 23  
 DB 1381 IGVPSANRRPELENNVIGFFVNTQICIRITIDENDNFESLVRQVRSRTTTAAQDNQDVPFEQ 1440  
 QY 24 ----- 23  
 DB 1441 VVSSLPSSSRDASRNPLVQLMFALHGQDLFKIQLEGTEEEVPTTEVTRFEDIEPHLYQ 1500  
 QY 24 ----- 23  
 DB 1501 GASKLSGDIIFAADLFEAETIRGVSVFQEVLRRLGLOQPPTPMTMPLTDGIPELERMGL 1560  
 QY 24 ----- 23  
 DB 1561 LHMVKTDPYRNMVVDVFOQQVRLSABEATVIDSSSRMSYASLDQSDQVAAWLRQLP 1620  
 QY 24 ----- 23  
 DB 1621 AETFFAVLAPRSCAEVIALFGILKAGHAYPLDVNVNPAARLRAILAELVKEKLVLLGAGE 1680  
 QY 24 ----- 23  
 DB 1681 PSPEQSPESVIVRIADATSPAGHASLRDGSKPTAGSLAYVIFTSGSTGKPKGVMEIHR 1740  
 QY 24 ----- 23  
 DB 1741 GVLRLVKQTNILSSLPAPQTFRMAHMSNLAFDASIEWEFTALLNGSLVICIDRFTILDQA 1800  
 QY 24 ----- 23  
 DB 1801 ALEALFLREHINIALPPPALLAQCLTDAATIKSLDILLVGGDRDLTADAALAKALVKSE 1860  
 QY 24 ----- 23



QY 98 -----GPI-QNYSITYLWF----- 110  
 Db 4081 VOQGSAPHPDPIVSTKYTGVPQSFQACRLWFLDQLNFGATWYLMPLAVLRGAMNVHALT 4140  
 QY 111 ----- 110  
 Db 4141 AALLALLERRHELLRTTFYBQNGVGMQVNPVVTETLRIIDLNSGDDGYLPTLKKEQTAPP 4200  
 QY 111 -----DFYSTQL----- 117  
 Db 4201 HLETEPGWRVALLRGLGPDYILSVVMHHIISDCWSVDVLFBQLGQFYSTAVKRGHDPLSQT 4260  
 QY 118 -----RXP----- 120  
 Db 4261 TPLPIHYRDPALWQKPKTOESEHEROLQYWVEQLVDSAPAEILLDLPFSLSGQAGMS 4320  
 QY 121 -----AKY----- 123  
 Db 4321 VTIEGALYKNLEBFCRVHRVTSFVLLAALRAAHYELTGSSEDATIGTPIANRRNPELEQI 4380  
 QY 124 ----- 123  
 Db 4381 IGFFVNTQCIRITVNEDETFSLVQOVRSTATAFAHQDVPEKIVSTLLPGSRDASRNP 4440  
 QY 124 -----VYSQVN----- 129  
 Db 4441 LVQLMPAVHSQKMLGELKLENASHSEVVPTEITTRFDLBFHLFQDDKLEGSILYSTDLFE 4500  
 QY 130 ----- 129  
 Db 4501 AVSVQSLLSVFOELRRGLNGPDVPISTLPQDGIVDLQROGLLDVQKTEYPRDSSVVDV 4560  
 QY 130 ----- 129  
 Db 4561 PHEQVSINPDSIALIHGSEKLSVAQLDRESDRVARWLHRSPSSDTLIAVLAPRSCETII 4620  
 QY 130 ----- 129  
 Db 4621 AFLGILKANLAYPLDVKAPAARIDAIVSLPGNKLILGANVTPPKLQEAADIFVPIRD 4680  
 QY 130 ----- 129  
 Db 4681 TPTTLTDGLQDGPHTIERPSAOSLAYAMFTSGTRPGKGMVQHRNIVRLVKNNSVNAKQ 4740  
 QY 130 ----- 129  
 Db 4741 PAAARIAHTSNLAFDASSWEIYAPLLNGGAIVCADYFTTIDPQALQETTFQEHIEIRGAMLP 4800  
 QY 130 ----- 129  
 Db 4801 PSLLKQCLVQAPDMISRLDILFAAGDRFSSVDALQARLVGSGVFNAYGPTENTILSTIY 4860  
 QY 130 ----- 129  
 Db 4861 NVAENDSFVNGVPIGSAVNSGAYIMDKNQQLVPAGVMGELVVTGDLARGYMDPKLDAD 4920  
 QY 130 -----HTAKTITFRP----- 139  
 Db 4921 RFIQLTVNGSEQVRAVYRTGDRVYRPQFQIEFFGEMDQOIKIRGHRIEPAEVEQAFLND 4980  
 QY 140 ----- 139  
 Db 4981 GFVEDVAIVIRTPENQEPWAFVTAKGNSAREEBEATTQIEGWEAHFEGGAYANIEETE 5040  
 QY 140 -----PPCGR----- 144  
 Db 5041 SEALGYDFMGWTSMYDGTIDKDEMRWLNDTMRSLLDGKPAGRVLEVTGTGTMIMFNILG 5100  
 QY 145 ----- 144  
 Db 5101 RSQGLERYIGLEPAPSAABFVNNAKSPFGLAGRAEVHVGTAAVGTGLOGLTSDMAVINS 5160

QY 145 -----VPSM----- 148  
 Db 5161 VAQYPTPEYLAETIKSLVQVFCMKRIYILGDMRSMWAMNRDFAAARAAYSILADNASKDRVR 5220  
 QY 149 -----TCLSEML----- 155  
 Db 5221 QKMMELEKEEBELLVDPAPFTALASQLQDRIQHVEILPKRMKATNELSSRYAAVLHISD 5280  
 QY 156 -----NVSKRNDTGE----- 165  
 Db 5281 EPLPIYKIDPEAWINFEGRSLTREALAQVLKENENAEVAISNIPYSKTVVERHIVRSLD 5340  
 QY 166 ----- 165  
 Db 5341 QEDANAPESMDGSDWISAVRTRAQOCHTSLASDLFDIAEDAGFRVEVSWARQHSQHAL 5400  
 QY 166 -----QG----- 167  
 Db 5401 DAVFHLKPATEDSRVLKFPPTDQGRPLKSLTNQPLLPQAQSRRAELLIREGLQTLPPY 5460  
 QY 168 -----CGNFT----- 172  
 Db 5461 MIPSQITLIDRMPLNANGKVDRRELARRAKITQSKPVEDIVPPRNSVEATVCKGFTDVL 5520  
 QY 173 -----TFNPMFNV----- 181  
 Db 5521 GVEVGITDN--FFNLGHSMLATKLAARLGRQLNTRISVRDVPDQPVVADLAAVIQNSA 5578  
 QY 182 ----- 181  
 Db 5579 PHEPIKPADYTGVPQSFQAGRLWFLDQLNVGATWYLMPLGIRLHGSLRVDALATAISAL 5638  
 QY 182 ----- 181  
 Db 5639 EQHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKDAYLAVLKHEQTTLFLDATE 5698  
 QY 182 PRWTKL----- 188  
 Db 5699 PGMRVALLIRLGBEEHILSIVMHHIISDCWSVEVLFDEMHRFYSSALRQDQPMELPLPI 5758  
 QY 189 ----- 188  
 Db 5759 QYRDPAAWQKTEQVAEHQRLQDYWTEHLADSTPAELLTDLPRPSILSGRANELPLTIEG 5818  
 QY 189 ----- 188  
 Db 5819 RLHDKLAPCRVHQATPFVILLALRAAHYRLTGAEDATLGTPIANRNRPELENMIGFFV 5878  
 QY 189 ----- 188  
 Db 5879 NTQCMRIAIBENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQVI 5938  
 QY 189 ----- 188  
 Db 5939 LAVHSQDGLKLTLEGLRDEAVDSASTRPFVFEHLFEHADRLSGSVLYAKELFKURTIE 5998  
 QY 189 ----- 188  
 Db 5999 SVSVFLETLRREALDQPLTFLAVLPLTDGVEIASKGLLDVPRTDYPRDANIVEVFQHV 6058  
 QY 189 ----- 188  
 Db 6059 RATPDIAVKAOTSILTYAQLDQSDRLAIVLSRRHMPETLVGLAPRSCETIIMFGI 6118  
 QY 189 ----- 188  
 Db 6119 MKANLAYPLDINSAPARLRSILSAVDGNKVLVLSGVTAPQENPEVEAVGIEIAGT 6178  
 QY 189 -----YV----- 190  
 Db 6179 GLDKTQGSNARPSATSLAYVIFTSSTGCKPGKGMVHRSVTRLAKPSNVISKLPOGARVA 6238  
 QY 191 ----- 190

Db	6239	HLANIAFDASIWEIATTLLNGATVCLDYHTVLDRTLKVEFERESITVVTLPALLKQC	6298
Qy	191	-----GPTKVNVDQSOTIY-----	203
Db	6299	VAEIPETLAHLDDLTYTGGDRVGGHDAMRARSLSVKIGMPSGYGPTENTVIS-TIYEVDADE	6357
Qy	204	-----	203
Db	6358	MFVNGVPIGKTVSNSGAYVMDRNOQLVPSGVVGLVVTGDGLARGYTDPSLNKNRFIYIT	6417
Qy	204	-----	203
Db	6418	VNGESIRAYRTGDRVRYRPHDLQIEFFGRMDQQVKIRGHRIEPEGESALLSHNSVQDAA	6477
Qy	204	-----	203
Db	6478	VVICAPADQSGABMVAFAVAARNTEDTQEBEAVDQVGWETHFETAAYSEVKDIRQSE	6537
Qy	204	-----FLGLTAL-----	210
Db	6538	VGNDFMGWTSMYDGSBIDKTMHEWLNDTMRMILDAREPGHVLIEIGTGTGMWFMFLAKCP	6597
Qy	211	-----	210
Db	6598	GLQYVGFEPSPKSAQAFVNDAAQSPFALKDGRSIVHVTATDINKAGPIQPRLVVINSVA	6657
Qy	211	-----	210
Db	6658	QYFPTPEYLFRVVEALVQIPSVVERIVFGDMRTNAINRDFVASRALHTLQKANKRLVRQM	6717
Qy	211	-----	210
Db	6718	IYELEANEELTDPAPFTSLRTRLGKIKHVBILPKTKMATNELSKYRYAAVLHVRGSR	6777
Qy	211	-----	210
Db	6778	EQSTIHQVSPNAMDFAADGLDRQTLINLLKEHKDAGTVAIGNIPYSKTIIVERFVNKSLS	6837
Qy	211	-----LLRYAQRNCT-----	220
Db	6838	EDDMEEGONSLDGSAMVAARVMAAQSPSLDAMDVKEIAQAGYQVEVSWARQWSQNGAL	6897
Qy	221	-----	220
Db	6898	DAIFHFPEPKEGARTLIEPPTDYEGRNVTLTNRPLNSTQSRRLGTQIREKLQTLILPEY	6957
Qy	221	-----	220
Db	6958	MIPSRIMVLDOMPVNNGKIDRKELVRRRAIVAPKPRSAATRVAPRNEIEAILRDEFEDVL	7017
Qy	221	-----HS-----	222
Db	7018	GTEVSVDNFFDLGGHSLMATKLAARVSRRLDAHISIKDVPDQVLADLAASIQRESAPH	7077
Qy	223	-----	222
Db	7078	EPIPORPYTGPAQSPAQGRMLFMDQLNLGATWYLMPLAIRINGQLRVAALSAALFALER	7137
Qy	223	-----	222
Db	7138	RHETLTTFEESDGVGVQIVGEARNSDLRVHDVSTGDDGEYLEVLREQTVPFDLSSEPG	7197
Qy	223	-----	222
Db	7198	WRVCLVKTGEEDHVLIVMHIIYDGVSDVILRGELGQFYSAALRGQDPLLHANPLPIQY	7257
Qy	223	-----	222
Db	7258	RDFAAWQREAKQVEEHQRGLGYSWKSQKOLVDSTPAELLTDLPRPSILSGRAGSVDTVIEGVS	7317
Qy	223	-----	222

Db	7318	YGALQSFRCRTRSVTTFVLLTVFRIAHFRLTAVDDATIGTPIANRNRPELETIVGCFVNT	7377
Qy	223	-----	222
Db	7378	QCMRISIADDDNPEGLVRQVRNVATAAYANQDVPFERIIVSALVPGSRNTRSNPLVQLMFA	7437
Qy	223	-----FVLV-----	226
Db	7438	VQSVEDYDQVRLEGLSVMPGCEASTRFDMEFHLVPGDKLGTSGVLYSSDLFEQGTQINF	7497
Qy	227	-----NAMSR-----NLFR-----	235
Db	7498	VDIFQCLASVLDQPLTPIVSLPFSNAISNLESLLDLEMTSDYPRDRTVVDLFRQAAI	7557
Qy	236	-----	235
Db	7558	CPDSIAVKDSSQLTYQAQDEQSDRVAWLHERHMPAESLVGLVSPRSCETIIAYFGIMK	7617
Qy	236	-----VPK-----	238
Db	7618	ANLAVPLDVYAPDARLAAILDTVSEGERLLLLGAGVPQPGIQIIPRLSTAYIAEALSHATT	7677
Qy	239	-----	238
Db	7678	VDVTIPOPSATSLAYVIFTSGTGKPGVMIHREGIVRLVRDVTNNVPFSGSALPVSH	7737
Qy	239	-----	238
Db	7738	FSNLAWDAATWEIYTAVLNGTVVICIDRDTMLDIAALNSTFRKENVRRAFFTPAFKQCL	7797
Qy	239	-----Y 239	
Db	7798	AETPELVANLEILTHTAGDRLDPGDANLAGTAKGGIFNVLGHTENTAYSTFFYPVVGEEFF	7857
Qy	240	ING-----	242
Db	7858	VNGVPGRGISNSHAYIIDRHQKLPAGVMGELIILTGDGVARGYTDSALNKDRFVYIDIN	7917
Qy	243	-----	242
Db	7918	GKSTWYRTGDKARYPRDQGLQEFFGRMDQVKIRGVRIEPEGEVELTLLDHKSVLAAATVV	7977
Qy	243	-----	242
Db	7978	VRRPPNGDPEMIAFITIDAEDDVQTHKAIYKHLQILPAYMIPSHLVILDOMPVTNGKV	8037
Qy	243	-----	242
Db	8038	DRKOLALRAQTQVKRRSTAAARVPPRDEVEAVLCEBYSNLLVEVGITDGFDFLGGHSLLA	8097
Qy	243	TKLK-----NT-----	248
Db	8098	TKLAARLSQLNTRVSVKXDFDQPIADLADIIRGSHRHDPIPATPYTGPVQSFQAGR	8157
Qy	249	-----	248
Db	8158	LWFLEQLNLGASWYLMFPAIRMGRPLQTKALAVNALVHRHEALRTTFEDHDGVGVQVI	8217
Qy	249	-----MRKLEKQKQ-----	257
Db	8218	QPKSSQDLRIIDLSDAVDDTAYLAALKREQTAFDLTSEPGRVSRLLRLGDDDDYILSIWM	8277
Qy	258	-----PVKEQ-----	262
Db	8278	HHIISDGTVDVLRQELGQFYSAAIRGQPELSQAKSLPIQYRDFAVMQRQENQIKEQAKQ	8337
Qy	263	-----	262
Db	8338	LKYWSQQLADSTPCFEFLDLPSPILSGEADAVPMVIGTVYQLLDFCETHQVTSFSL	8397
Qy	263	-----PEKRAK 269	
Db	8398	LAAFRTAHLRTGTLDATVGTPIANRNRPELEGLIGFFVNTQCMRMAISETETESLVOQ	8457

QY	270	TOSTTTPFS-----	279	QY	304	-----	303
Db	8458	VRLTTTEAFANQDPPEQIVSTILLPGSRDTSRPLVQVMFALQSQQDLGRIOLEGMTDEA	8517	Db	9538	ARRAQVAKRKAVSARVAPRNDTEIVLCBEYADILGTEVGITDNFFDMGSHLMATKLA	9597
QY	280	-----	279	QY	304	-----	303
Db	8518	LETPLSTRLDLVHLPQEVGKLSGSLYSTDLFEVETIRGIVDVFLIILRGLGEPQKRL	8577	Db	9598	RLSRRLDTRVTVEKVPDKVLADLAASIEQGSTPHLPIASSVYSGPVEQSYAQGRWLWFLD	9657
QY	280	-----	279	QY	304	-----	303
Db	8578	MAMPITDGITKLARDQGLLTVAKPAYPRESSVIDLFRQOVAAPDAFAVWDSSTLIYADL	8637	Db	9658	QPNLNAATWYHMSLAMRLGPLNMDALDVALRALEQRHETLRTTFFBAQKDIGVQVVHEAGM	9717
QY	280	-----	279	QY	304	-----	303
Db	8638	DGQSNKLAHWLCORNMAPETLVAFPAPRSLCTIVAFGLVKANLAYLPDLDVNAAPARIEA	8697	Db	9718	KRLKVLDSLCKNEKEHMAVLENEQMRPPTLASEPGWKHGLARLGPTETIYLSLVMHMFSD	9777
QY	280	-----	279	QY	304	-----	303
Db	8698	IILSAVEGHKLVQAHGPELGLTMAOTELVQIDEALASSSGDHQIHASGPTATSLAYV	8757	Db	9778	GWSVDILRQELGQFYSAALRGDRDPLSQVKPLPIQYRDFPAWQKEAAQVAEHERQLAYWEN	9837
QY	280	-----	279	QY	304	-----	303
Db	8758	METSGTGKPGVMIDHRSIIRLVKNSDVVATLPTPRMANVSNLAFDISVQEIYTALIN	8817	Db	9838	QLADSTPGELLTDFPRPQFLSGKAGVIVPTIEGPPVEKLLKFSKERQVTLFVSLLTAPRA	9897
QY	280	-----	279	QY	304	-----	STST 307
Db	8818	CGTLVCLDYLTLLDSKILYNVFEAQVNAAMFTVLLKQCLGNMPAIIISRLSVLFNVGDR	8877	Db	9898	THFRLTGAEDATIGTPIANRNRPELEHIIIGFFVNTQCMRLLLDTGSTPESLVQHVRSVAT	9957
QY	280	-----	279	QY	308	IAY-----	310
Db	8878	LDAHDAVASGLIQDAVYNAYGPTENGMOSTMVYKVDVNEPFVNGVPIGRSITNSGAYMD	8937	Db	9958	DAYSNQDIPFERIVSALLPGSRDASRPLIQMPALHSQPDIGNITLEGLEHERLPTSVA	10017
QY	280	-----	284	QY	311	-----	310
Db	8938	GNQQLVSPGVMEIVVTGDLARGYTDLSALDEDRFVHVITDGEENIKAYRTGDRVRYRPK	8997	Db	10018	TFDMEFHLFQEPNKLSGSILLFADELFPQETINSVVTVFQETLRRGLDQPVSI	10077
QY	285	-----	284	QY	311	-----	310
Db	8998	DPEIEBFGMDQVKIRGHRIBPAEVEHALLGHDLVHDAAVVLRKPANQEPSEMIAPITQ	9057	Db	10078	DGLIDLEKLGLEIBSSNFPDRDYSVVDVFRQOVAANPNAPAVVDSETSMYS	10137
QY	285	-----	284	QY	311	-----	314
Db	9058	EDETIEQHSNKQVQGEHFDVSRYADIKDLDTSTFGHDFLGWTSMDYGVDPVNMKE	9117	Db	10138	IAWLHAQGLRPESLICVWAPRSFETIVSLFGILKAGYAYLPDVSNPAARIQPI	10197
QY	285	-----	284	QY	315	-----	314
Db	9118	WLDETTASLLDNRPQGHILEIGAGTGWILSNLKGVDGLQYVGLDPAPSAALFVNEAVKS	9177	Db	10198	GKRLVLLSGIDMPQSDRMDVETARIQDILLTNKVERSDDPMSRPSATSLAYVIFT	10257
QY	285	-----	284	QY	315	-----	314
Db	9178	LPSLAGKARVLVGTALDIGSLDKNEIQPELAVINSVAQYFPTSEYLIKVVKA	9237	Db	10258	RPKGVNIEHRNITLVLKQSNVTSQLPQDLRMAHISNLAFAASIWIIFTAILNGG	10317
QY	285	-----	287	QY	315	-----	318
Db	9238	RVFFGDIRSQALNRDFLAARAVRALGDNASKEQIREKIAELBESEELLVDPAFFVSLRS	9297	Db	10318	YFTLLDSQALRTTFEKARVNATLFPALLKECLNAPTLPEDLVKVIYIGDRDL	10377
QY	288	-----	287	QY	319	-----	318
Db	9298	QLPNIKHVEVLPKMKATNELSSRYAAVLHISHNEEBQLLIQDIPTAWVDF	9357	Db	10378	IQALVKGTVYNAYGPTENTVMSTIYRLTDGESYANGVPIGNVSSGAVIMDKQ	10437
QY	288	-----	303	QY	319	-----	318
Db	9358	QGLRNLLQGRDDVMIAVGNIPYSKTIIVERHIMNSLDQDHVNSLDGTSWIS	9417	Db	10438	GVWGLVVGDCGLARGYTNSTLNADRFVDIVINDQKARAYRTGDRTRVPK	10497
QY	304	-----	303	QY	319	-----	318
Db	9418	TSFPALATLAKEGFRVELSWARQSONGALDAVFHRLATDANCERSRVLVHP	9477	Db	10498	NDQQVKIRGHRVEPAEVBQAMLGNKAIHDAVVVQAVDQGETEMIGFVSMAS	10557
QY	304	-----	303	QY	319	-----	318
Db	9478	GRQLRTLNRPLQRAQSRRIESQVFEALQATALPAYMIPSRIIIVLPQMP	9537	Db	10558	EITNOQWEDHFEFESTAYATGIEAIDQATLGRDFTSWTSMYNGNLIDKAEME	10617
Db				QY	319	-----	318

Db	10618	LLDKEDARPCAEIGTGTGMVLENLPKNDGLESYVGIEPSSAALFVDKAAQDPFGLQGT	10677	Db	11698	QAVKLHANSVRFTTRISDALVESGSPTEBELSTRPTAQSLAYVMFTSGSTGPKGVNVEHR	11757
QY	319	-----	318	QY	346	-----	345
Db	10678	QILVGTABDIKLVKDFHPDWWVINSVAQYFSPRSYLQIASELIHMTSVKTIFFGDMRSW	10737	Db	11758	GITRLVKNNSVNAKOPAAAAIAHLSNIAFDASSWEIYAPLLNGTVCIDYTTIDIKAL	11817
QY	319	-----	318	QY	346	-----	345
Db	10738	ATNRDFLVSRALYTLGDKATKDQIRQEVARLEENEDELLVDPAFTSLTSQWPKVKHVE	10797	Db	11818	EAVFKQHHIRGAMLPALLKQCLVSAPTMISSLEILFAAGDRLLSSQDAILARRAVSGVY	11877
QY	319	-----	318	QY	346	-----	345
Db	10798	ILPKRMETSNELSSRYAAVLHICRDGEGRNRYGRRVHSVEENAWIDFASSGMDRHVLQ	10857	Db	11878	NAYGPTENTVLSTIHNIGENAEFSGVPICGNAVSNSGAFVMDQNOQLVSAGVIGELVWTG	11937
QY	319	-----	318	QY	346	-----	345
Db	10858	MLDERRDAKTAIGNIPHSNTINERHFTTSLDTGEGIAQDSLDGSAMOSATKAMAARCP	10917	Db	11938	DGLARGYTDSKLURDRFIYITLDGNRVRAYRTGDRVRHRPKDQIEFFGMDQOQIKIRGH	11997
QY	319	-----	318	QY	346	-----	345
Db	10918	CLSVTELVEIGQAAGPRVEVSUARQSRQHGALDVVFHHEDDRGRVLINFTDFELPP	10977	Db	11998	RIEPAEVEQALARDPAISDSAVITQLTDEEPELVAFPSLKGNGANGTNGVGSDOEKID	12057
QY	319	-----	318	QY	346	-----	353
Db	10978	STGLTSRPLQRIQNRFPESQIREQLQTLPPYVMPVSRIVVLERMPLNANSKVDRKELARK	11037	Db	12058	GDEQHALLMENKIRHNLAQALLPTMIPSRIIHVDQLPVNANGKIDRNLAVRAQAATPTS	12117
QY	319	-----	325	QY	354	-----	356
Db	11038	ARTLQTIKPSATVAPNDIEAVLCDFEQAVLGVTVGVMDNPFPELGGHSLMATKLAARLS	11097	Db	12118	SVSTVAPNDIETIICKEFADILSVRGITDNFDFLGLGHSIATKLAARLRRDTRVS	12177
QY	326	-----	325	QY	357	-----	356
Db	11098	RRLDTRVSVKDIENQPIQLADVVQTSAPHEAIPSTYSGPVEQSFQGRMLFIDLQNL	11157	Db	12178	VRDVFDPVVGQLAASIQQGSTPHEAIPALSHSGPVQSQFAGRLWFLDRFNLNAAWYTM	12237
QY	326	-----	325	QY	357	-----	365
Db	11158	LNASWYHMLASRLRGLRIEALQSALATIEARHESLRTTFEBQDGPVQIVRAARNKQL	11217	Db	12238	PFVRLRGLRVDALQALRALAEERHELLRTTFEBQDGVGMQIVHSPMRDICVVDISGA	12297
QY	326	-----	325	QY	366	NE-	367
Db	11218	RIIDVSGTEDAYLAALKQBODAAFDTAEFGWVALLRLGPDHVLISVMHHIISDGWSV	11277	Db	12298	NEDLAKLEEQAQPNLSTEVANRVALFKAGENHHILSIVMHHIISDGWSVDIFQOELAQ	12357
QY	326	-----	331	QY	368	-----	367
Db	11278	DILRQELGQLYSNASSQAPLPPIQYRDFAIWQDQSIABHOKQLNWKQLVNSKPAEL	11337	Db	12358	FYSVAVRGHDPLSVQVKPLPIHYRDFAVMQRQKVAVHESQLQYWIEQLADSTPAEILSD	12417
QY	332	-----	339	QY	368	-----	367
Db	11338	LADFTPKALSGDADVIPIEIDDQVYONLRSFCRARHVTFSVALLAFAAHYRLTGABD	11397	Db	12418	FNRPEVLSGEAGTVPIVIEDEVYEKLSLFCRNHQVTSFVLLAAPRAHYRLTGADATI	12477
QY	340	-----	339	QY	368	-----	367
Db	11398	ATIGSPIANRPELEGLICGFVNTQCLRIPIVKS EDTFTLVKQARETATEAQDNQDVFF	11457	Db	12478	GTFPIANRNRPELEDLIGFFVNTQCMRIALAEHDNPLSVVRRVRSTAASAFENQDVFFERL	12537
QY	340	-----	343	QY	368	-----	371
Db	11458	ERIVSSWASSRDTSRNPLVQVMFAVSHQDLGNIRLEGVEGKPVSMASATRFDAEMHLF	11517	Db	12538	VSALLPGSRDASRNPLVQLMFMVHVSQRNLGKLEGEGETPYTATTTFDVFPHLEFD	12597
QY	344	-----	343	QY	372	-----	371
Db	11518	EDQGM/LGNNVFSKOLFSEETIRSVVAVFQETLRRGLANPHANLATPLTDGLPSLRSLC	11577	Db	12598	KGLAGNVVFAADLFEAATIRSVVEVFHEILRRGLDQPDIAISTMPLVDGLAALNSRNLP	12657
QY	344	-----	343	QY	372	-----	377
Db	11578	LQVNPDPYRDASVIDVPREQVASIPKSTAVIDASSQLTYTTELDRSSQLATWLRQVTV	11637	Db	12658	VEDIEPDPATEASVVDVFTQVAVNPDALAVTDTSTKLTLYAELDQSDHVAWLSKQKLP	12717
QY	344	-----	345	QY	378	-----	377
Db	11638	PBELVGLAPRSCETIIAFGLIIKANLAYPLDVNAPAGRIETILSSLPGNRLILGSDT	11697	Db	12718	AESIVVVLAPRSSETIVACIGILKANLAYLPMDSNVPEARQAIISEIPGKFKVLLGAGV	12777
QY	346	-----	345	QY	378	-----	377
				Db	12778	PIPNKTDADVRMVFISDIVASKTKYSYSGTRPSASSLAYVIFTSGSTGRPKGVNVEHRG	12837

QY 378 ----- 377  
Db 12838 VISLVKQNASRIPQSLRMAHVSNAFADASVWEIFTTILANGTILFCISYFTVLDSKALSAA 12897  
QY 378 -----MSSLY-----YN-- 384  
Db 12898 FSDHRINITLLPPALLKQCLADAPSVLSLESYIGGDRLDGADATKVKDLVKGKAYNAY 12957  
QY 385 ----- 384  
Db 12958 GPTENSVMSTIYTIETHPANGVPIGTSGLGPKSKAYIMDQDQQLVPAGVMGELVAGDGL 13017  
QY 385 -----E 385  
Db 13018 ARGYTDPSLMTGRFIHITIDGKQVAYRTGDRVYRPRDYQIEFFGRDQOIKIRHRIE 13077  
QY 386 TMEVENKTASDS----- 397  
Db 13078 PAVEQALLSDSSINDAVVSAQNKEGLEVMGVITTOAAQSVDKBEASNKVQWEAHFDS 13137  
QY 398 ----- 397  
Db 13138 TAYANTIGIDRDALGQDFLSWTSWYDGLIPREMQEWLNDTWRSLLDNQPPGKVLEIGT 13197  
QY 398 -----NKTTPSPM----- 407  
Db 13198 GTGMVLFNLGKVBGLQSYAGLEPSRSVTAVNKAIBTFPSLAGSARVHVGTAEIDISSDG 13257  
QY 408 -----GFQRTF----- 413  
Db 13258 LRSDLVINSVAQYFSPREYLAELTANLRLPGVKRIFFGDMRTYATNKDFLVARAVHTL 13317  
QY 414 ----- 413  
Db 13318 GSNASKAMVRQVAKLEDDDEELLVDPAFTSLSDQPPDEIKHVEILPKRMAATNELSSY 13377  
QY 414 ----- 413  
Db 13378 RYAAVHVGHGHPNGEDEDKQAVKDINPKAMVDFAGTMDRQALLQLLQDRQRDDVV 13437  
QY 414 ----- 413  
Db 13438 AVSNIPYSKTIEMRHLQSLLDDEDTGSAVDGTAMISRTQSRACEPALSVADLIBIGK 13497  
QY 414 ----- 413  
Db 13498 IGFEVEASWARQHSQGGILDVPHRFEPPRPHSHGVMPFTEHKGRSSSLTNRPPLHLIQ 13557  
QY 414 ----- 413  
Db 13558 SRRLEAKVRERLQSLPPYMPISRTITLQDQPLTSNGKVDKRLARQARVIPSAASTLD 13617  
QY 414 ----- 413  
Db 13618 FVAPRTEIEVLCEEFTDLLGVKVGITDNFFELGSHSLATKLSARLSRRLDAGITVKQV 13677  
QY 414 -----IDPLM----- 418  
Db 13678 FDQPVLAADLAASILQSSRHSRPSLPGYGPVEQSPAQGRNLWFLDQFNIDALWYLPFAL 13737  
QY 419 ----- 418  
Db 13738 RMGPQLQVDALAAALVALEERHESLTTFEERDGVGIQVVQPLRTTKDIRIIDVSGMRDD 13797  
QY 419 ----- 418  
Db 13798 DAYLEPQEQOTPFPLASEPGWRVALLKLGKDDHILSIWMHIIISDGNSTEVLORELQ 13857  
QY 419 -----D 419  
Db 13858 FYLAASKGKAPUSQVAPLPIQVRDPAVQORQEQVAESQRLDYWKQLADSPAPELLAD 13917

QY 420 Y----- 420  
Db 13918 YTRPNVLSGEAGSVSFVINDSVYKSLVFCRSRQVTTFTTLLAAFRAAHYMTGSDDATI 13977  
QY 421 ----- 420  
Db 13978 GTPIANRPELENLIGCFWNTQCMRITIGDDTFESLVQOVRSTTATAPENQDVPPERI 14037  
QY 421 -----LDSLLF----- 426  
Db 14038 VSTLSAGSRDTSRNPLVQLLFAVHSQQGLRIQLDGVWDEPVLSTVSTRFDLSEHAFQEA 14097  
QY 427 -----LDEI----- 430  
Db 14098 DRLNGSMVFATDLFQPETTIQGFVAVVEEVLQRLGEQFQSPMATPLAEGIAQLRDAGALQ 14157  
QY 431 -----RNFSL----- 435  
Db 14158 MPKSDYPRNASLVDFVFOQQAAMASPVAVTVDSTSKUTYAE LDLRSDQAASYLRRQQLPAE 14217  
QY 436 -----RS----- 437  
Db 14218 TWAVLAPRSCETIIAFLAILKANLAYMPLDVNTPSARMEALISSVPGRRLLVSGVRH 14277  
QY 438 ----- 437  
Db 14278 ADINVPNAKTMLTISDTVTGTAIGTPEPLVVRPSATSLAYVIFTSGTGPKGMVEHRA 14337  
QY 438 ----- 437  
Db 14338 IMRLVKDSNVTHMPPATRMAHVNTIAFVSLFEMCATLLNGTLVICIDYLTLLDSTMRL 14397  
QY 438 ----- 437  
Db 14398 ETFEREQVPAALFPALLRQCLVNMPPDAIGMLEVYVAGDRPHSRDARATQALLAGPRVYN 14457  
QY 438 -----PT-----YN-----LTPP----- 446  
Db 14458 AVGPTENAILSTIYNIDKHDYVNGVPIGSAVNSGAYVMDRNQQLLPGVMGELVVTGE 14517  
QY 447 ----- 446  
Db 14518 GVARGYTDASLDTDRFVTVIDGQRQRAYRTGDRVYRPGFQIEFFGRDQOAKIRGHR 14577  
QY 447 -----EH----- 448  
Db 14578 VELGEVEHALLSENSVTDAAVLRMTMEEDPQLVAFVTTDHEVYRSGSSNEEDPYATQAA 14637  
QY 449 -----RAVNLSSTNS----- 459  
Db 14638 GDMRKRLRSLLPYMYVPSRVITLQPLNANGVKDRKLARRAQMTPTASSSGPVHVAPR 14697  
QY 460 ----- 459  
Db 14698 NETEAAICDBFETILGVKVGITDNFFELGSHSLATKLAARLSRMLRISVKDLFDDPV 14757  
QY 460 ----- 459  
Db 14758 PVSLLAKLEQQQGFSGDESSTVGIVPFQLLPAEMSREIIQRDVVPOIENGHSTPLDMYP 14817  
QY 460 ----- 459  
Db 14818 ATQTQIFFLHDKATGHATPPPLSLDFPETADCRRLASACAALVQHFDIFRTVFSRGGR 14877  
QY 460 ----- 459  
Db 14878 FYQVVLHLDVPEVETEQEDELVALHEADKQOPLRGLRAMLRILKRPKAKORLV 14937  
QY 460 -----LM----- 461  
Db 14938 LRMSHSLYDGLSLEHIVNALHALYSDKHLAQAPKFGLYMHMHASRRRAGYNFWRSLQGS 14997  
QY 462 -----W----- 462

Db	14998	SMTSLKRSVGALEAMTPSAGTWTQTSKSIIRIPPAALKNGITQATLFTAAVSLLAKHTKST	15057
QY	463	-----W 463	462
Db	15058	DVVFGRVSGRQLSINCQDIVGPCINEVPRVRIDEGDDMGLLRAIQDYTSRPHET	15117
QY	463	-----W 463	
Db	15118	LGLQEVKENCTDW	15130
RESULT 9			
ADP84155	ADP84155 standard; protein; 22157 AA.		
XX	AC	ADP84155;	
XX	DT	26-AUG-2004 (first entry)	
XX	XX	Human CA125 protein sequence SeqID 5.	
DE	XX	human; CA125; antigenic determinant; ovarian carcinoma cell;	
KW	XX	serum tumour marker; ovarian cancer; cytostatic.	
KW	OS	Homo sapiens.	
OS	XX	Key	
XX	XX	Location/Qualifiers	
FT	XX	Misc-difference 1. :22152	
FT	XX	/label= Xaa	
FT	XX	/note= "Xaa can be any naturally occurring amino acid"	
XX	XX	WO2004045553-A2.	
PN	XX	03-JUN-2004.	
PD	XX	17-NOV-2003; 2003WO-US037041.	
PF	XX	15-NOV-2002; 2002US-0427045P.	
PR	XX	(UYAR-) UNIV ARKANSAS.	
XX	XX	O'brien T, Beard J, Underwood L;	
XX	XX	WPI; 2004-420533/39.	
DR	XX	N-PSDB; ADP84151, ADP84152, ADP84153, ADP84154.	
XX	XX	New isolated nucleic acid molecule encoding CA125 protein, useful for	
PT	XX	diagnosing or treating cancer, particularly ovarian carcinoma.	
PT	XX	Claim 14; SEQ ID NO 5; 501pp; English.	
PS	XX	This invention relates to a novel nucleic acid molecule identified as	
XX	XX	CA125, and the encoded protein thereof. Specifically, it refers to the	
CC	XX	cloning, identification and expression of the CA125 gene, and in	
CC	XX	particular the glycosylated amino terminal domain, multiple repeat domain	
CC	XX	and the carboxy terminal domain. The present invention describes CA125 as	
CC	XX	an antigenic determinant that is expressed on ovarian carcinoma cells,	
CC	XX	yet show essentially no expression in normal adult ovarian tissue.	
CC	XX	Accordingly, this protein represents a serum tumour marker that can be	
CC	XX	used for early diagnosis and improved prognosis in cancer patients,	
CC	XX	specifically ovarian cancer, such that they can be used to develop	
CC	XX	pharmaceutical compositions that exhibit cytostatic activities. This	
CC	XX	polypeptide sequence is the human CA125 protein of the invention.	
XX	XX	Sequence 22157 AA;	
SQ	XX	Query Match 22.9%; Score 561; DB 8; Length 22157;	
	XX	Best Local Similarity 1.2%; Pred. No. 0.0027;	
	XX	Matches 249; Conservative 77; Mismatches 137; Indels 20726; Gaps 80;	
QY	XX	2 GRKEMVRDVPKMFVLIS-----	19

Db	755	GTPEISTTKPSSAVLSSMTLSNAATSPERVRNATSPLTHPSPSGEETAGSVLTLSAE	814
QY	20	-----	19
Db	815	TTDSPNIHPTGTLTSSSESPSTLSLPVSGVKVTFSSSTPSTHLPTSGETEETSNPVS	874
QY	20	-----	19
Db	875	SQPETSVSRRVTTLASTSVPTPVFPTMDTWPTRSAQFSSSHLVSELRASTSVTNSTGS	934
QY	20	-----	19
Db	935	ALPKISHLTGATMSTQNRDTFNDAAPOSITTWTPETSPRFTGLPSATTVTSTATLSA	994
QY	20	-----	19
Db	995	TVVSKFTSPATYSMEATSIREPSTTILTTETNGPGSMNAVASTNIPICKGYITEGRDLT	1054
QY	20	-----	19
Db	1055	SHLPIGTTASSETSMDFTWAKESVMSVSPSQSMDAAGSSTPGRTSQFVDTPFDDVVHLT	1114
QY	20	-----	19
Db	1115	SREITIPRDGTSSALTPQMTATHPPSPDPFGSARSTWLGLSSSPSSPTPKVTMSSTFSTQ	1174
QY	20	-----	19
Db	1175	RVTTSIMDVTETSRWNPNLSTTSLTSPSNIPTSGAICKSTLVPLDTPSPATSEASEG	1234
QY	20	-----	19
Db	1235	GLPTLSTYPESTNTPSIHLGAHASSESPSTINLTWASVVKPGSYTLTPFSIETHIVST	1294
QY	20	-----	19
Db	1295	ARMAYSSGSSPEMTAPGETNCGTWDPPTYITTPDKOTSQAQVSTPHSVRTLRTTENHP	1354
QY	20	-----	19
Db	1355	KTESATPAAYSGSPKISSPNLTSPATKAWTITDTEHSTQLHYTKLAKSSGFQTSAP	1414
QY	20	-----	19
Db	1415	GPVSVVPIPTSPITGSGTLELTSVDPGEPLVLAPSEQTITLPMATWLSTLSEMASTDL	1474
QY	20	-----	19
Db	1475	DISPSSPMSTFAIPFPWSTPSHELKSEADTSAIRNTDSTTLDQHLGIRSLGRTGDLTT	1534
QY	20	-----	19
Db	1535	VPITPLTTTWSVIEHSTQADTSLSATMSPHVTQSLKDOTSIPASAPSHLTVVPELG	1594
QY	20	-----	19
Db	1595	TQGRSSSEATFWKPSDTLRSREIETGPTNIQSTPMDNTTTCSSSSGVTLGIAHLPIGT	1654
QY	20	-----	19
Db	1655	SSPAETSTNMALERRSSTATVSMAGTMGLLVTSAPGRSISQSLGRVSSVLSSSTTEGVTD	1714
QY	20	-----ISPL-----	23
Db	1715	SSKGSPLNTQGNLTALSSSLEPSVABGQMSTSIPLTSSPTPDVEFIGGSTFWTKVT	1774
QY	24	-----	23
Db	1775	TWMTSDISKSSARTESSATLMSTALGSTENTCKEKLRTASMDLPSPSPSMEVTPWISLT	1834
QY	24	-----	23
Db	1835	LSNAPNTTDSLDSHGVTSSAGTLATDRSLNTGVTTRASRLNGSGDTSKSLSMGNSTHT	1894



QY	24	-----	39	-----	38
Db	1895	SMTDEKSEVSSIHPRPETSAPGAETTLTSTPGNRAISLTLPLFPSSIPVEEVISTGITSG	2975	EBRVPTSGTGDPRYASESMSYPDPKASSAMTSTSLASKLTLTFLSTGQAARSGSSSPIS	3034
QY	24	-----	39	-----	38
Db	1955	PDINSAPMTHSPITPTTIWVTSTGTIEQSTOPLHAVSEKSVQSTOPYVNSVAVASP	3035	LSTEKETSLPSTASTSRKTSILPLGSMARQPNILVHLQTSALTSTLSTLNMSQEBPPE	3094
QY	24	-----	39	-----	38
Db	2015	THENSVSSGSTSTPYSSASLESLSDTISRRNAITSLMLDLTSLTPTTTPSTLSSEALS	3095	LTSSQTIABEEGTAAETQTLTFTPTSETPTSLLPVSSPTEPTARRKSSPETWASSISVPAK	3154
QY	24	-----	39	-----	38
Db	2075	SGHSGVSNPSTTTTEPPLFSAASTSAKORNPETETHGPQNTAASTLNTDASSVTGLSET	3155	TSLVETTDGTLVTTIKMSQAAQGNSTWPAABETGTSAGTSPGSPGVSTTLKIMSSKE	3214
QY	24	-----	39	-----	43
Db	2135	PVGASISSEVPLPMAITSRSDVSGLTSESTANPSLGTASAGTKLRTTISLPTSESLVSF	3215	PSISPEIRSTVNSPWKTPETTVPMETTVETLQSTALGSGSTSISHLPTGTTSTPKSP	3274
QY	28	-----	44	-----	43
Db	2195	RMNKDPWTVSILPLGSHPTTNTETSIIPVNSAGPGLSTVASDVIDTPSDGARSPTVVSFP	3275	TENMLATERVSLSPSPPEAWTNLYSGTPOGTRQSLATMSSVSLESPTARSITGTGQSSP	3334
QY	28	-----	44	-----	43
Db	2255	SPOTEVTTIHPEKTHSFRITISLTHELTSRVTRIPGDMSSAMSTKPTGASPSITLG	3335	ELVSKTTGMEFSGMHGSGTGGTTHVLSLSTSSNILEDPTVSPNSVSSLTDKSKHKKTETW	3394
QY	28	-----	44	-----	43
Db	2315	ERRITSAAPTSPIVLTASTETSTVSLDNETTVKTSILDARKTNELPSDSSSSDLI	3395	VSTTAIPSTVLNKKIMAAEQTSRVSDEAYSSSTSSWSDGTSDDITLGASPDVNTLYIT	3454
QY	29	NCKVM-----	44	-----	43
Db	2375	NTSIASSTMDVTKASISPTISGMTASSPSLFSDDRPOQVPTSTTETWTATSPSVSNT	3455	STAQTSLVSLPSGQGITSLNPSGGKTSSASSVTSPSIGLETLRANVSAVKSDIAPTA	3514
QY	34	-----	44	-----	43
Db	2435	YSLDGSNVGGTPSTLPPFTIHPVETSSALLAWSRPVTFSTWVSTDGTASGENPTSSNS	3515	GHLQSOTSPAEVSIILDVTTAPTFGISITTIITMTGNSISTTTPNPEVGMGTMDSTPATER	3574
QY	34	-----	44	-----	43
Db	2495	VVTSVPAGTWASVGSTTDLPAWGFLKTPAGEAHSLLASTIEPATAFPHLSAAVVTS	3575	TTSTBHPWMSSTAASDSWTVTDMTNSLNKVARSPGTISTMHTTSPFLASSTELDSMTPHG	3634
QY	34	-----	44	-----	43
Db	2555	SATSEASLLTTSEKAIHSSPQPTPTTSGANWETSATPESLLVVTETSDTLTSLKILVT	3635	RITVIGTSLVTPSSDASAVKTETSTISERTLSFSDTTASTPTSTFSRVQRMISVDPILST	3694
QY	39	-----	44	-----	43
Db	2615	DTILFSTVSTPPSKFPSTGTLGASFPILLPDTPAIPLTATEPTSSLATSPDSTPLAVTIA	3695	SWTPSSTEADVPVSMVPTDHASTKTDPNTPLSTFLFDSLSTLDMDTGRSLSSATATSA	3754
QY	39	-----	44	-----	43
Db	2675	SDSLGTVPETTLTMTSETNGDALVLKTVSNPDRSIPGITIQGVTEPLHPSSTSPSKIVA	3755	POGATTPOBELTLETWISPATSPQLPFSIGHTSAVTPAAMARSSGVTFSPDPTSKAEQT	3814
QY	39	-----	44	-----	43
Db	2735	PRNTVEGSI VALSTLPAGTTGSLVFSOSSENSETTALVDSSAGLERASVMPLTGSOQ	3815	STQLPTTTSAPHCQVPRSAATLTDVIPHAKTPDATFQOQTALTTEARATSDSWNEKE	3874
QY	39	-----	44	-----	43
Db	2795	MASSGGIRSGTHSGTKTFSSLPLTMNPGEVTAMSEITNRLTATQSTAPKGI PVKPTS	3875	KSTPSAPWITEMMNSVSEDTIKEVTSSSVLKDPEYAGHKLGIWDDFI PKFGKAAHREL	3934
QY	39	-----	44	-----	43
Db	2855	ABSGLLTPVSASSPSKAFASLTAPPSTWGI PQSTLTTFEFSEVPSLDTKSASLPTPGOS	3935	PILLSPPQDKEAIHPSTNTVETTTGWVTSSEHSHSTIPAHASASSKLTPVVTSTREQAIV	3994
QY	39	-----	44	-----	43
Db	2915	LNTIPDSDASTASSLSKSPKPNRARMMTSTKAISASSFQSTGTETPETGASPSMAGH	3995	SMSTTTWPESTARTEPNSFLTIELRDVSPYMDTSTTQTSIISSPGSTAIYKGRTEIT	4054
QY	39	-----	44	-----	43

Db	4055	SYKRISSEFLAQMRRSDSPSEAITRLSNFPAMTESGGMILAMQTSPPGATSISAPTLD	4114	Db	5135	SGHESHSPALADSETPKATTQWVITTTVGDPAPTSMPPVHGSSETNIKREPTFLTPRL	5194
QY	44	-----	43	QY	47	-----LSKI-	50
Db	4115	SATASWTGTLATTQRTYSEKTLTLPKGREDTSPSPPCVBEETSSSSVVPIHATTSPS	4174	Db	5195	RETSTSQSSPPDTSFLLSKVPTGTITEVSTGVISSSKISTDPHDKSTVPPDTFTGEI	5254
QY	44	-----	43	QY	51	-----	50
Db	4175	NILLTSQGHSPSTPPTVTSVFLSETSGLGKTTDMSRI SLEPGTSLPPNLSSTAGEALSTY	4234	Db	5255	PRVFTSSIKTKSAEMTITTOASPPESASHSTLPLDTSTLSQGGTHSTVSQGPYSEVTT	5314
QY	44	-----	43	QY	51	-----	50
Db	4235	EASRDTKAIHSGADTAVTNMEATSEYSPIPGHTKPSKATSPLVTHIMGDITSSTSVFG	4294	Db	5315	LMGMPGNVSMWMTTTPVEETSSVSLMSSPAMTSPSPVSSTSPPQIPSPSPLPVLTALPTSV	5374
QY	44	-----	43	QY	51	-----	50
Db	4295	SSETTEIETVSSVNOQLQERSTSOVASSATETSTVITHVSSGDATHVTKTQATPSSGTS	4354	Db	5375	LVTITDVLGTTSPESVTSSPPNLSIITHERPATYKDTAHTAAMHHSNTAVTNVGTSGS	5434
QY	44	-----	43	QY	51	-----	50
Db	4355	IGSPHQFITSINTFTDVSTNPSTSLIMTESSGVITITQTGPTGAATQGPYLLDTSMPYL	4414	Db	5435	GHSQSSVLADSETSKATPLMSTASTLGDTSVSTPNISQTNQIQTEPTASLSPRLRES	5494
QY	44	-----	43	QY	51	-----	50
Db	4415	TETPLAVTDFMQSEKTLTISKPKDVTWTSPSVAETSYPSLTPFLVTTIPPATSTLQ	4474	Db	5495	STSEKTSSTTETNTAFSVPTGCAITQASRTEISSRSTISDLDRSTIAPDISGCMITRLF	5554
QY	44	-----	43	QY	51	-----	50
Db	4475	GQHTSPVSATSVLTSGLVKTTDMLNTSMBPVTNSPONLNPNSEILATLAATTDIETIH	4534	Db	5555	TSPIMTKSAEMTVTTQTTTPGATSQGILPMDTSTTLFOGGTHSTVSQGPHEIITLRSR	5614
QY	44	-----	43	QY	51	-----	50
Db	4535	PSINKAVTNMGTAASSAHVLHSTLPVSSEPTSTATSPMVPASSMGDALASISIPGSETTOIE	4594	Db	5615	TPGDVSMWMTTTPVEETSSGFSLMSPMTSPSPVSSTSPEISPLPVLTALLTSVLVTT	5674
QY	44	-----	43	QY	51	-----	50
Db	4595	GEPTSSLTAGKENSTLOEMNSTESNIIILSNVSGAITEATKMEVPSDATFIPTPAQS	4654	Db	5675	NVLGTTSPPEVTSPPNLSSTQERLTTYKDTAHTAMHASMHTTAVANVGTSGHES	5734
QY	44	-----	43	QY	51	-----	50
Db	4655	TKFPDIFSVASSRLSNSPMTISTHMTTQTGSSGATSKIPLALDSTLTETAGTSPSVT	4714	Db	5735	QSSVPADSHTSKATSPMGITFAMGDTSVVTSPTAFPETRIQSESTSSLIPGLDRTRTSEE	5794
QY	44	-----	43	QY	51	-----	50
Db	4715	EGFAHSKITTAMNDVKDVSQTNPPQDEASSPSSQAPVLVTTLPSSVAFTPQWHSTSSP	4774	Db	5795	INTVTETSVLSEVPTTTTTEVSRTEVITSSRTTISGPDHSMSPYISTETITRLSTFPF	5854
QY	44	-----	43	QY	51	-----	50
Db	4775	VSMSSVLTSGLVKTAGKVDTSLETVTSPOQSMNTLDDISVTSAAATDITTHPSINTVV	4834	Db	5855	VTGSTEMAITNQTFIGITISQATLTLDTSSTASWEGTHSPVTQRPHESETTMSRSTKG	5914
QY	44	-----	43	QY	51	-----	50
Db	4835	TNVGTTGSAFESHSTVSAPEPSKVTSNVTTSTMEDTIIISRISIPKSKTTRTETETSS	4894	Db	5915	VSMQSPPSVEETSSPSPVPLPAITSHSLSYSAVSGSSPTSALPVTSLTSGRRKTIDML	5974
QY	44	-----	43	QY	51	-----	50
Db	4895	LTPKLRSTSISOEITSSSTETSTVPYKELTGATTEVSRKTDVTSSSSTSPGPDQSTVSLDI	4954	Db	5975	DTHSELVTSLSLPASASSFSGEILTSEASTNTETIHFSENTAETNMGTNSMHLHSSVSIH	6034
QY	44	-----	43	QY	51	-----	50
Db	4955	STETNRLTSPIMTESABITITQTQPHGATSQDFTTWDPNNTTPQAGIHSAMTHGFSQ	5014	Db	6035	SQPSCHTTPPKVTGSMEDAIIVSTSTPGSPETKNVDRDSTSLTPELKEDSTALVMNSTTE	6094
QY	44	-----	43	QY	51	-----	50
Db	5015	LDVTTILMSRIPODVSWTSPPSVDKTSPPSFLSSPAMTTPSLISSTLPEDKLSSPMTSLL	5074	Db	6095	SNTVFSSVSLDAATEVSRABEVTYDPTFMPASAQTKGPDISPESASSHSNSPPLTISTH	6154
QY	44	-----	46	QY	51	-----GKYKLDQKLE-	61
Db	5075	TSGLVKITDILKTRLEPVTSSLNPNFSGTSDKILATSKDSKDTKEIPPSINTEETNVKANN	5134	Db	6155	KTIAITQGPBGVTSGLQTLDTSTIATISAGTSPARTQDFVDSETTSMNNDLNDVLKTSR	6214
QY	47	-----	46	QY	62	-----	61
				Db	6215	FSABEANSLSQAPLLVTTTSPSVTSTLQEHSTSLSVSVTPTTLAKITDMDTNLEPV	6274

QY	62	-----	98	-----	97
Db	6275	TRSPQNLRLTATSEATTDTHMHPINTAMANGTSSPNEFYFTVSPDSDPYKATSAV	7355	ITHSTMTQRLPHSEITTLVSRGAGDVPRPSPPLVEETSPSSQLSLSAMISPPVSSTLPL	7414
QY	62	-----	98	-----	97
Db	6335	VITSTSGDSIVTSMRPSAMKIESETTFSLIPRLRETSTSQKIGSSSDTSTVPDKAFT	7415	ASSHSSASVTSPLTPGQVKTEVLDAEPTSPPPSLSSSTSVEILATSEVTTDTTEKIH	7474
QY	62	-----	98	-----	97
Db	6395	AATTEVSRTELTSRSSRTSIQTEKPTMSPDTSRSTVMTSLFAGLTKEERTIAQTGPH	7475	PPNTAVTKVGTSSSGHESPSPVLPDSETTKATSAMGTISIMGDTSVSTLPALSNTRKI	7534
QY	62	-----	98	-----	97
Db	6455	RATSOGLTWDTSITTSQAGTHSAMTHGFSQDLSTLTSRVPEYISGTSPPSVEKTS	7535	QSEPASSLTURLRETSTSEETSLATEANTVLSKVGTGATTEVSRTEAISFSRTSMGPGQ	7594
QY	62	-----	98	-----	105
Db	6515	SLLSLPAITSPSPVPTLPESRPSSPVHLTSLPTSGLVKTTDMLASVASLPPNLGTS	7595	STMSQDISIGTIPRISASSVLTSASAKMTITTTQGPSESTLESTLNLNTATTFSWVETHSI	7654
QY	62	-----	106	-----	105
Db	6575	IPTTSEDIKDTEKMPSTNIATNVGTTTSEKESYSSVPAYSEPCKVTSFPMVTSNIRDT	7655	VIQGFPHPEMTTSMGRGPGVSWPSPFPVKETSPSSPLSLPAVTSHPVSTFLAHI	7714
QY	62	-----	106	-----	105
Db	6635	IVSTSMPSGSEITRIEMESTFSLHGLKGTSTSQDPVSTEKSAVLHLKLTTCATET	7715	SPLPVTLTSGPATTTDILGTSTBPTGTSSSSLSTTSHERLTYKYDTAHTBAVHPST	7774
QY	67	-----	106	-----	105
Db	6695	VASSRRTSIPGPDHSTESPDISTEIVPSLPIGLITESNNMTIIRTGPPLGSTSQGT	7775	GGTNVATTSSGYKSOSSVLADSSPMCTTSTMGDTSVLTSTPAFLETRRIQTELASLT	7834
QY	67	-----	106	-----	110
Db	6755	LDTPTSSRAGTHSMATQBFPHSEMTVMNKDPEILSWTIPPSIEKTSFSSSLMPS	7835	LRSSGSECTSGTKMSTVLSKVPTGATTEISKEDVTSIPGPAQTSISPDTRT	7894
QY	67	-----	111	-----	110
Db	6815	SPPVSTLTKTIHTTPSMTSLTLPVLMTDTLGTSPBPTSSPDLSTSHVILTDE	7895	TSPVMTSEAITMNTHTSPLGATTQGTSTLDTSTSTSLTMTHTSTISQGFHQSM	7954
QY	67	-----	111	-----	110
Db	6875	DTTAIEAMHPSTATAINVETTCSGHGSQSVLTDSEKTKATAPMDTSTMGHTT	7955	GPEDVSWSPPLLEKTRFSPSLMSSPATTSPPVSPSTLPESISSPPLVTSLLT	8014
QY	75	NVSKQ-----	111	-----	110
Db	6935	SVSSETTKIKRESTYSLTPLRLTSIQNASFSDTSIVLSEVPTGTTAEVSRTE	8015	TDMLHKSSEPVTNSPANLSSSTSVEILATSEVTTDTEKTHPSSNRRTVDVG	8074
QY	80	-----	111	-----	110
Db	6995	RTSIPGQSQTVLPEISTRMTRLPASPTMTBSAEMTIPTQTGPGSGTSDTLT	8075	FVLADQTSKVTPMVITSTMEDTSVSTSTPGFFETSRITQETPSLTGLRKT	8134
QY	80	-----	111	-----	110
Db	7055	KSQAKTHSTLQRFPHSEMTLMSRPGDMQSQSPSLENPSLPLSLPATTSP	8135	SLATEMTVLGCVPTGATAEVSRTVETSSRTSISGFAQLTVSPETSTETIRL	8194
QY	80	-----	111	-----	110
Db	7115	STLPVTISSPLVTSLLTSSPVTMTDMLHTSPELVTSPPKLSHTSDE	8195	TESAEWMIKTQDPPGSTPESHTVDISTTPNWEVTHSTVTQRFHSEMTL	8254
QY	89	E-----	111	-----	110
Db	7175	EAVHPSTNTAASNVBIPSGHESPSSALADSETSKATSPMFTSQDITVAIS	8255	LWPSQSVETSSASSLLSLPATTSPSPVSTLVEDFPASLPVTSLLTPGLVIT	8314
QY	90	-----	111	-----	110
Db	7235	TSRIQKESISSLPKLREGSSVETSSAIEITSAVLSEVIGATTEISRTVTS	8315	ISREPGTSTSNLSTSHERLTTLEDVDTAMQPSHTAVTNVRTSISGHESQ	8374
QY	90	-----	111	-----	116
Db	7295	GSAESTMLPEISTTRKIKRFPSPILABSEMTIKTQTSPPGSTSESTTLT	8375	SETPKATSMGTYTMMGETSVSISTDFETSRVQIEPTSSLTSGLRETS	8434
QY	90	-----	117	-----	116

Db	8435	GSTVLSEVPSGATTEVSRTEVISSRGTSMSGPDQFTISPDISTEAITRLSTPIMTESAE	8494	Db	9515	HVPENTAASHVETTSATEELYSSSPGFSEPTKVLSPVVTSSIRDNMVSTMPGSSGTR	9574
QY	117	-----	116	QY	131	-----	130
Db	8495	SAITITGSPGATSEGTLLDTSTTTFMGTHSTASPGFHSHEMTTLMSTRPGDVPWPSL	8554	Db	9575	IEIESMSLTPGLRETRTSQDITSSTETSTVLYKMSSGATPEVSRTEWMPSSRTSIPQPA	9634
QY	117	-----	116	QY	131	-----	130
Db	8555	PSVEERASSVSSLSPPAMTSTSPFASALPESISSPHPVTAITLPGVKTTDMLRTSSPE	8614	Db	9635	QSTMSLDISDEVVTRLSTSPIMTESABEITTTQTYSLATSQVTLPLGTSMFLSGTHST	9694
QY	117	-----	116	QY	131	-----	130
Db	8615	TSSPPNLSSTSAEILATSEVTKOREKIHPSSNTPVVNVGTVIYKHLSPSSVLADLVTKP	8674	Db	9695	MSQGLSHSEMTNLMRGPESLSWTSPPREVTRSSSSILTSPLTSLSPVSSILLDSSPS	9754
QY	117	-----	116	QY	131	-----	130
Db	8675	TSPMATTSTLGNTSVSTPAPFETMMQTSSLTSGLEISTSQETSATERSASLSGM	8734	Db	9755	SPLPVTSLILPLGVKTTVEVLDTSSEPKTSSPNLSSTSVETPATSEIMTDEKIHPSNT	9814
QY	117	-----	116	QY	131	-----	130
Db	8735	PGATTKVSRTEALSIGRTSTPGPAQSTISPEISTETITRISTPLTTGSAEMTTPKTG	8794	Db	9815	AVAKVRTSSVHSHSVLADSETTITIPSMGITSAVDDTTVTNSPAPSETRRIPEPT	9874
QY	117	-----	116	QY	131	-----	130
Db	8795	HSGASSQGTFTLDTSSRASWPGTHSAATHRSPHSGMTTPMSRGPEDEVSWPSPSVEKTSP	8854	Db	9875	FSLTPGPRETSTSEBTTSITETSAVLYGVPTSATTEVSMTEIMSSNRTHIPDSQSTMS	9934
QY	117	-----	116	QY	131	-----	140
Db	8855	PSSLVLSAVTSPSPLYSTPSSSHSPLRVTSLFTFPVMMKTTDMLDTSLBVPVTTSPPM	8914	Db	9935	DIITEVITRLSSSSMMSESTQMTTITQKSSPGATAQSTLTATLATTAPLARTHSTVPPRFL	9994
QY	117	-----	122	QY	141	-----	140
Db	8915	NITDESLSATKATMETEALQLSENTAVTQMGTISARQEFYSSYPGLPEPSKVTSPPVTS	8974	Db	9995	HSEMTLMSRSPENPSWKSPFVEKTSSSSSLLSLPVTTSVSVSTLPQSIPSSSFSVTS	10054
QY	123	-----	122	QY	141	-----	140
Db	8975	STIKDIVSTTIPASSEIRIEMESTSTLPTPRETSTQEIHSATKXPSTVPYKALTSATI	9034	Db	10055	LLTPGMVKTDTSTEPGTSLSPLNSLGTSTVEILAASEVTTDEKIHPSSSMAVTVNGTSS	10114
QY	123	-----	122	QY	141	-----	143
Db	9035	EDSMTQVMSSSRGSPDQSTMSODISTEIVITRLSTSPIKAESTEMTITQTGSPGATSRG	9094	Db	10115	GHELYSSVSIHSEPSKATYPVGTPTSSMAETSISTSMANPETTGFEAPFHLTSGFRKT	10174
QY	123	-----	122	QY	144	-----	143
Db	9095	TLTLDSTTFMSGTHSTASQSPSHSQMTALMSRTPGDVPWLHSPVEEASASFSLSPPV	9154	Db	10175	NMSLDTSVTNTPTSPSPGTHLQSKTDTSSAKTSSPDWPPASQYTEIPVDIITPEN	10234
QY	123	-----	122	QY	144	-----	143
Db	9155	MTSSSPVSTLPDSIHSSSLPVTSLTSLGLVKTTELLGTSSEPETSPPNLSSTSAELA	9214	Db	10235	ASPSITESTGITSPESPESRFTMSVTESTHLSLTDLLPSAETISTGTVMPSLSEAMTSFAT	10294
QY	123	-----	130	QY	144	-----	143
Db	9215	TTEVTDTKELEMTNVVTSYTHESPPSVLADSVTTKATSSMGITYPTGDTNVLSTPAF	9274	Db	10295	GVPRASGSGSPFSRSTESPGDATALSTIAESLPSTPVPFSSSTFTTSDSTIPALHEIT	10354
QY	131	-----	130	QY	144	-----	143
Db	9275	SDTSRIOTKSLSTPGLMETSISEETSSATEKSTVLSVPTGATTEVSRTEAISSRST	9334	Db	10355	SSSATPYRVDTSLGTESSTEGRLVNVSTLDTSSQPGRTSSTPILDTMTESVELGTVTS	10414
QY	131	-----	130	QY	144	-----	148
Db	9335	IPGPAQSTMSDTSMETITRISTPLTRKSTDMAITPKTGPAGATSGGTFTLDSSTASW	9394	Db	10415	AYQVPSLSTRLTRTDGIMEHITKI PNEAAHRTGIRPVKQPSTSPASPGLHTGKTGM	10474
QY	131	-----	130	QY	149	-----	155
Db	9395	PGTHSATQRPQSVVTPMSPRGPEDEVSWPSPLEKNSPPSSILVSSSVTSPLYSTP	9454	Db	10475	ETTTTALKTTTALKTTSRATLTTSVYTPTLGLTFLNASROMASTILTEMMITPYVPP	10534
QY	131	-----	130	QY	156	-----	155
Db	9455	SGSSHSPVPVTSFTSIMMKATDMLDASLEPETTSAPNMNITDESLSATSKATTETAI	9514	Db	10535	DVPETSSLSATSLGAETSTALPRTTSPVNLRSSETTASLVSRSGAERSPVIQTLDVSSSE	10594
QY	131	-----	130	QY	156	-----	159
				Db	10595	PDTTASWVIHPAETIPTVSKTTPNFHSELDTVSSSTATSHGADVSSAIPNTNISPELDAL	10654

```
QY 160 ----- 159
Db 10655 TPLVTSIGTDTSTFFTLTKSPHETETRTTTLWTHPAETSSITPRIPNFSHIESDATPSI 10714
QY 160 ----- 161
Db 10715 ATSPGAETSSAIPIMTVSPGAEDLVTSQVTSSTGDRNMTPTLTLSPGEPKTIASLVTHP 10774
QY 162 ----- 161
Db 10775 EAQTSSAIPSTISPAVRLVTSMTVSLAAKSTTNBALTNSPCEPATTVSLVTHPAQTS 10834
QY 162 ----- 161
Db 10835 PTVPWTTISFFHKSDDTTPMTTSHGAESSAVPTPTVTEVPGVTVPLVTSRAVISST 10894
QY 162 ----- 161
Db 10895 IPILTLSPGEPETTPSMATSHGEBEASSAIPPTVSPGVGVTVSLVTSRAVTSITPIL 10954
QY 162 ----- 161
Db 10955 TFSLGEBETTPSMATSHGTEAGSAVPTVLPVPGWTVSLVASSRAVTSITLTLSPGE 11014
QY 162 ----- 161
Db 11015 PETTPSMATSHGAESAVPTVSPVPGVTVSLVTSSSGVNSTIPTLILSPGELETTPS 11074
QY 162 ----- 161
Db 11075 MATSHGAESSAVPTPTVSPGVGVTVPLVTSRAVTSITPILTLSSSEPETTPSMATS 11134
QY 162 ----- 161
Db 11135 HGVEASSAVLTVSPVPGWTVSLVTSRAVTSITPILTLSSDEPETTVSLVTHSEAKMI 11194
QY 162 ----- 161
Db 11195 SAIPTLAVSPVQGLVTVSLVTSSTGSETSAFNLTVASSQETIDSWAHPGTEASSVPT 11254
QY 162 ----- 165
Db 11255 LTVSTGEPTFNISLVTHPAESSLPRTRTSRFGHSELDTMPSTVSPAEASSAISTTIS 11314
QY 166 ----- 165
Db 11315 PGIPGVLTSLVTSSTGRDISATFTVPVSPESPHESSEATASWTHPAVTSITVPRTPPNYSHSE 11374
QY 166 ----- 165
Db 11375 PDTPSIATSPGAESATDPPTITVSPDVPDMVTSQVTSSTGDTSTIPTLTLSSGPEIT 11434
QY 166 ----- 174
Db 11435 TSFITYETHSSAIPPLVPSPGASKMLATSLVSSGDTSTTPTTLTETPYEBETTAIOL 11494
QY 175 ----- 178
Db 11495 IHPAETNTWVPRTPPAFSHKSDDTLPLVAITSPGPEASSAVSTTISPDMSDLVTSLVPS 11554
QY 179 ----- 178
Db 11555 SGTDTSTPTTLSETPYEBETTATLWTHPAETSTTVSGTIPNFSHRGSDTAPSMVTSVPG 11614
QY 179 ----- 179
Db 11615 DTRSGVPTTIPPSIPGVVTSQVTSATDTSTAIPTLTPSGPEPETTASSATHPGTQGF 11674
QY 180 NVP----- 182
Db 11675 TVPIRTVPSSEPDTMASWTHPPQTPTVPSRTTSSFSHSSPDATPVNATSPREASSAVL 11734
QY 183 ----- 182
Db 11735 TTISPGAPENVTSQITSSGAATSTTVPTLTHSPMPETTALLSHTRPTETSKTFFPASTVF 11794
QY 183 ----- 182
Db 11795 PQVSETTASLTIRPGAETSTALPTQTTSLSFTLLVGTSRVDLSPTASPGVSAKTAPLST 11854
QY 183 ----- 182
Db 11855 HPGTETSTMIPTSTLSLGLLETTGLLATSSAETSTSTLTLTVSPAUGLSSASITTDKP 11914
QY 183 ----- 187
Db 11915 QVTSMNTETSPSVTVSGPPEFSRTVTGTTMLIPSEMPPTPKTSHGEGVSPVTILRTTM 11974
QY 188 ----- 187
Db 11975 VEATNLATGSSPTVAKTTTTFNTLAGSLFTPLTTPGMSTLASESVTSRTSYNHRSWIST 12034
QY 188 ----- 187
Db 12035 TSSYNRRYVTPATSTPTVSTFSPGISTSSIPSSSTAATVPWVPFTLNTITNLQYEDMR 12094
QY 188 ----- 192
Db 12095 HPGSRKFNATERLOGLLKLPRNSLELYLSCRLASLRPEKSSAMAVDAICTHRPDP 12154
QY 193 TKVNDVDSQTIYP----- 204
Db 12155 EDGLDRERLYWELSNLTNGIOELGPYTLDRNSLYNGFTHRSSMPTTTPGTSTVDVGT 12214
QY 205 ----- 204
Db 12215 SCTPSSSPPTAAGPLLMPFTLNTITNLQYEDMRRTGSRKENTMESVLQGLLKLFPKN 12274
QY 205 ----- 204
Db 12275 TSVGLYSGRLTLRLRPEKGAATGVDAICTHRLDPKSPGLNREQLYWELSKLNDIEEL 12334
QY 205 ----- 204
Db 12335 GPYTLDRNSLYNGFTHQSSVSTTTPGTSTVDLRTSGTPSSLSPTIMAAGPLLVPTL 12394
QY 205 ----- 204
Db 12395 NFTITNLQYGEDMCHPGSRKFNTERVLOGLLGPFNKTSVGLYSGCRLTSLSRSEKGA 12454
QY 205 ----- 204
Db 12455 ATGVDAICIHHLDPKSPGLNRRLYWELSQLTNGIKELGPYTLDRNSLYNGFTHRTSVP 12514
QY 205 ----- 216
Db 12515 TSSTPGTSTVDLGTSGTPSPSPATAGPLLVLTFLNTITNLKYBEDMHRPGSRKFNNT 12574
QY 217 ----- 216
Db 12575 ERVLQTLGPMFNKTSVGLLYSGCRLTLRSEKGAATGVDAICTHRLDPKSPGLDREQL 12634
QY 217 ----- 216
Db 12635 YWELSQLTNGIKELGPYTLDRNSLYNGFTHWIPVPTSSTPGTSTVDLGSCTPSSLPST 12694
QY 217 ----- 216
Db 12695 AAGPLLVPTLNTITNLQYEDMCHPGSRKFNTERVLOGLLGPFNKTSVGLLYSGCR 12754
QY 217 ----- 218
Db 12755 LTLRSEKGAATGVDAICTHRLDPKSPGVDRQLYWELSQLTNGIKELGPYTLDRNSLY 12814
QY 219 ----- 218
```

Db	12815	VNGFTHQTSAPNTSTPGTSTVDLGTSGTSPSLPSTPSAGPLLVPTFLNFTITNLQVEEDM	12874	Db	13895	LGPYXLDXSLYVNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXXTSAGPLLVPTFLN	13954
QY	219	-----CTH---	221	QY	253	-----	252
Db	12875	RHPGRKFNTTERRVLOGLLKPLFKSTSVGLYSGCRLTLRLRSEKGAATGVDAICTHRLD	12934	Db	13955	FTITNLQVEEDMHHPGSRKFNTTERRVLOGLLGPMPFKNTSVGLLYSGCRLTLRLRPEKGAA	14014
QY	222	-----	221	QY	253	-----	252
Db	12935	PKSPGVDRQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLG	12994	Db	14015	TGMDAICSHRLDPKSPGLDRQLYWELSQLTHGIKELGPYTLDRNSLYVNGFTHRSSVAP	14074
QY	222	-----SFVLVN-----	227	QY	253	-----KPK-----	255
Db	12995	TSCTPSSLPSPSTAGPLLVPTFLNFTITNLQVEEDMHHPGSRKFNTTERRVLOGLLGPMPFK	13054	Db	14075	TSTPGTSTVDLGTSGTSPSSLPSPPTAVPLLVPTFLNFTITNLQYEDMRHHPGSRKFNTTE	14134
QY	228	-----	227	QY	256	-----QAP-----	258
Db	13055	NTSVGLLYSGCRLTLRLRPEKNGAATGMDAICSHRLDPKSPGLNRQLYWELSQLTHGIKE	13114	Db	14135	RVLQGLLGPLFKNSSVGLYSGCRLISLRSEKGAATGVDAICTHHLNPQSPGLDREQLY	14194
QY	228	-----AMSRN-----	232	QY	259	-----VKE-----	261
Db	13115	LGPYTLDRNSLYVNGFTHRSSVAPTSTPGTSTVDLGTSGTSPSSLPSPPTAVPLLVPTFLN	13174	Db	14195	WQLSOMTNGIKELGPYTLDRNSLYVNGFTHRSSGLTSTPWTSTVDLGTSGTSPSPSPPT	14254
QY	233	-----LFR-----	235	QY	262	-----	261
Db	13175	FTITNLQYEDMRHHPGSRKFNTTERRVLOGLLGPLFKNSSVGLYSGCRLISLRSEKGA	13234	Db	14255	TAGPLLVPTFLNFTITNLQYEDMRHHPGSRKFENATERVLOGLLSPIFKNSSVGLYSGCR	14314
QY	236	-----VPK-----	235	QY	262	-----	261
Db	13235	TGVDAICTHHLNPQSPGLDREQLYWLSQMTNGIKELGPYTLDRNSLYVNGFTHRSSGLT	13294	Db	14315	LTSLRPEKGAATGMDAVCLYHPNPKRPGLDREQLYWELSQLTHNITELGPYSLDRDLSL	14374
QY	236	-----	238	QY	262	-----QPEKKA	267
Db	13295	TSTPWTSTVDLGTSGTSPSPPTAGPLLVPTFLNFTITNLQYEDMRHHPGSRKFENATE	13354	Db	14375	VNGFTHQSSMTTRTPDSTWHLATSRTPASLSGTTASPLLVLTINCTITNLQYEDM	14434
QY	239	-----	238	QY	268	KKTQS-----	272
Db	13355	RVLQGLLSPIFKNSSVGLYSGCRLTSLRPEKGAATGMDAVCLYHPNPKRPGLDREQLY	13414	Db	14435	RTGSRKFNTMESVLOGLLKPLFKNTSVGLYSGCRLTLRLPKKGAATGVDAICTHRLD	14494
QY	239	-----	238	QY	273	-----	272
Db	13415	WELSQLTHNITELGPYSLDRDLSLYVNGFTHQNSVPTSTPGTSTVYVWATGTPSPPGHT	13474	Db	14495	PKSPGLNRQLYWELSQLTNDIEELGPYTLDRNSLYVNGFTHQSSVSTTSTPGTSTVDLR	14554
QY	239	-----	238	QY	273	-----	272
Db	13475	EPGPLLPFTFNFTITNLHYEENQHPGSRKFNTTERRVLOGLLKPLFKNTSVGLYSGCR	13534	Db	14555	TSGTPSSLSPTIMXXXPLLPFTFLNFTITNLXYEEXXPGSRKFNTTERRVLOGLLRPL	14614
QY	239	-----Y	239	QY	273	-----	272
Db	13535	LTSLRPEKGAATGMDAVCLYHPNPKRPGLDREQLYCELSQLTHNITELGPYSLDRDLSL	13594	Db	14615	FKNTSVSSLYSGCRLTLRLRPEKGAATRVDAACTVYRDPDPKSPGLDREQLYWELSQLTHSI	14674
QY	240	ING-----	242	QY	273	-----	272
Db	13595	VNGFTHQNSVPTSTPGTSTVYVWATGTPSSFPCHTEPGPLLPFTFNFTITNLHYEENM	13654	Db	14675	TELGPYTLDRVSLYVNGFNPRSSVPTTSTPGTSTVHLATSGTPSSLPGHTXXXPLLPPT	14734
QY	243	-----	242	QY	273	-----	272
Db	13655	QHPGSRKFNTTERRVLOGLLKPLFKNTSVGLYSGCRLTLRLRPEKHEAATGVDTICTHRVD	13714	Db	14735	LNFTITNLXYEEXXPGSRKFNTTERRVLOGLLKPLFRNSSLEYLYSGCRLASLRPEKDS	14794
QY	243	-----TKLQNTMRKL-----	252	QY	273	-----	272
Db	13715	PIGPGLDRERLYWELSQLTNSITELGPYTLDRDLSLYVNGFNPRSSVPTTSTPGTSTVHLA	13774	Db	14795	SAMAVDAICTHRPDPEDLGLDRERLYWELSQLTNGIQELGPYTLDRNSLYVNGFTHRSSF	14854
QY	253	-----	252	QY	273	-----TTTTPYFS-----	279
Db	13775	TSGTPSSLPCHTAPVPLLPFTFLNFTITNLHYEENQHPGSRKFNTTERRVLOGLLKPLFK	13834	Db	14855	LTTSTPWTSTVDLGTSGTSPSPPTAGPLLVPTFLNFTITNLQYEDMRHHPGSRKFNT	14914
QY	253	-----	252	QY	280	-----	279
Db	13835	NTSVGLYSGCRLTLRLRPEKHEAATGVDTICTHRVDPIGPGLDREXLYWELSQLTXIXE	13894	Db	14915	TERVLOGLLTPLFKNTSVGLYSGCRLTLRLRPEKQBAATGVDTICTHRVDPPIGPGLDRR	14974
QY	253	-----	252	QY	280	-----	279
Db	13895	LYWELSQLTNSITELGPYTLDRDLSLYVNGFNWSSVPTTSTPGTSTVHLATSGTPSSLP	15034	Db	14975	LYWELSQLTNSITELGPYTLDRDLSLYVNGFNWSSVPTTSTPGTSTVHLATSGTPSSLP	15034

```
QY 280 ----- 279
Db 15035 HTAPVLLIPFTLNFTITDLHYEENMHQPGSRKFNFTERVLOGLLPLFKSTSVGLYSY 15094
QY 280 -----YT----- 281
Db 15095 CRLTLRPEKHGAATGVDAICTLRDPTGGLDRERYWELSQLTNSVTELGPYTLDRDS 15154
QY 282 -----TSA-----ALNVT--TNVTY-- 294
Db 15155 LYVNGFTHRSSVPTTSIPGTSVAHLETSGTPASLPGHTAPGLLVPFTLNFTITNLQYEE 15214
QY 295 ----- 294
Db 15215 DMRHPSGRKFTTERVLOGLLPLFKNTSVSSLYSGCRLTLRPEKGAATRVDAVCTHR 15274
QY 295 -----SITTAARVSTSI- 308
Db 15275 PDPKSPGLDRERYWKLSQLTHGTELGPYTLDRHSLYVNGFTHQSSMTTTRPDTSTMH 15334
QY 309 ----- 308
Db 15335 LATSRTASLPGTTASPLLVLTINFTITNQRYEENMHQPGSRKFNFTERVLOGLLRPV 15394
QY 309 -----AYRPDS----- 315
Db 15395 FKNTSVGLYSGCRLTLRPPKDKGAATKVDAICTYRPDPKSPGLDREQLYWELSQLTHSI 15454
QY 316 ----- 315
Db 15455 TELGPYTQDRDSLYVNGFTHRSSVPTTSIPGTSVAHLETSGTPASLPGHTAPGLVPT 15514
QY 316 ----- 315
Db 15515 LNFTITNLQYEEBMRHPSGRKFNFTERVLOGLLPLFKSTSVGLYSGCRLTLRPEKRG 15574
QY 316 ----- 315
Db 15575 AATGVDTICTHRLDPLNPLDREQLYWELSKLTRGIELGPYLLDRGSLYVNGFTHRTSV 15634
QY 316 ----- 315
Db 15635 PTTSTPGTSTVDLGTSGTPSLPSPAXXXPLXPFNLFTITNLXYEEXXPGSRKFNFT 15694
QY 316 ----- 315
Db 15695 TERVLOTLLGPMFKNTSVGLLYSGCRLTLRSEKGAATGVDAICTHRLDPKSPGVDRQ 15754
QY 316 ----- 315
Db 15755 LYWELSQLTNGIKELGPYTLDRNSLYVNGFTHWIPVPTSTPGTSTVDLGTSGTPSLPSSP 15814
QY 316 -----FMKSIMATQUR----- 326
Db 15815 TTAGPLLVPFTLNFTITNLKYEBDMHCPGSRKFNFTERVLOGLLPMFKNTSVGLYSGC 15874
QY 327 ----- 326
Db 15875 RLTLRSEKGAATGVDAICTHRLDPKSPGVDRQYWELSQLTNGIKELGPYTLDRNSL 15934
QY 327 -----DLAT----- 330
Db 15935 YVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSLPSPPTXXXPLXPFNLFTITNLXYEEX 15994
QY 331 ----- 330
Db 15995 MXXPGSRKFNFTERVLOGLLPLKFKXKTSVGLYSGCRLTLRXXEKKAAATXVDXXCXXX 16054
QY 331 -----WV----- 332
Db 16055 DPXXPGLDREXLYWELSKLTXIXELGPYXLDKXSLYVNGFTHWIPVPTSTPGTSTVDL 16114
QY 333 -----YTTLRVRQNPFCBPS----- 347
Db 16115 GSGTPSSLPSPTTAGPLLVPFTLNFTITNLKYEBDMHCPGSRKFNFTERVLOGLLPMFK 16174
QY 348 ----- 347
Db 16175 NTSVGLYSGCRLTLRSEKGAATGVDAICTHRLDPKSPGVDRQYWELSQLTNGIKE 16234
QY 348 -----RNRTAVSEF----- 356
Db 16235 LGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPSPTSAGPLLVPFTLN 16294
QY 357 -----MKNTHV----- 362
Db 16295 FTITNLQYEBDMHPSGRKFNFTERVLOGLLPMFKNTSVGLLYSGCRLTLRPEKNGAA 16354
QY 363 ----- 362
Db 16355 TGMDAICTHRLDPKSPGLDREXLYWELSKLTXIXELGPYXLDKXSLYVNGFXXXXXXX 16414
QY 363 ----- 362
Db 16415 TSTPGTSVXLXTSGTPXXXPKXXXPLXPLXPFNLFTITNLKYEEXXPGSRKFNFT 16474
QY 363 -----LIRNET----- 368
Db 16475 RVLOGLLPLFRNSSLEVLYSGCRLASLPEKDSAMAVDAICTHRPDPEDLGLDRERY 16534
QY 369 -----PVTI----- 372
Db 16535 WELSNLNGIQELGPYTLDRNSLYVNGFTHRSMPTTSTPGTSTVDVGTSGTPSSPSPT 16594
QY 373 -----YG----- 374
Db 16595 TAGPLLIPTLNFTITNLQYGEDMHPGSRKFNFTERVLOGLLGPFKNTSVGLYSGCR 16654
QY 375 -----TLDMSLY 382
Db 16655 LTSRSEKGAATGVDAICIIHLDPKSPGLNRERYWELSQLTNGIKELGPYTLDRNSLY 16714
QY 383 YN----- 384
Db 16715 VNGFTHRTSVPTSTPGTSTVDLGTSGTPSPATAGPLLVLFTLNFTITNLKYEDM 16774
QY 385 ----- 384
Db 16775 HRPGRKFNFTERVLOTLLGPMFKNTSVGLLYSGCRLTLRSEKGAATGVDAICTHRLD 16834
QY 385 ----- 384
Db 16835 PKSPGLDREXLYWELSKLTXIXELGPYXLDKXSLYVNGFXXXXXXXSTPGTSXVXLX 16894
QY 385 ----- 384
Db 16895 TSGTPXXXPKXTXXXPLXPLXPFNLFTITNLXYEEXXPGSRKFNFTERVLOGLLRPVK 16954
QY 385 ----- 384
Db 16955 NTSVGLYSGCRLTLRPPKDKGAATKVDAICTYRPDPKSPGLDREQLYWELSQLTHSITE 17014
QY 385 -----ETMFVENKTASDSNKT----- 401
Db 17015 LGPYTQDRDSLYVNGFTHRSSVPTTSIPGTSVAHLETGTPSPFPGHTBPGLLIPTFN 17074
QY 402 ----- 401
Db 17075 FTITNLRYEENMHQPGSRKFNFTERVLOGLLPLFKNTSVGLYSGCRLTLRPEKQAAA 17134
QY 402 ----- 401
Db 17135 TGVDTICTHRLDPIGGLDRERYWELSQLTNSITELGPYTLDRDSLYVDGPNPSSVPT 17194
QY 402 ----- 401
```

Db	17195	TSTPGTSTVHLATSGTPSLPGHTAPVPLLIPTFLNFTITDLHYEENMOHPSRKFNTE	17254	Db	18275	VNGFTHQNSVPTTSTPGTSTVYWAATGTPTSSPFGHTPGPLLIPTFTNFTITNLHYEENM	18334
QY	402	-----PTSPSMG-----	408	QY	432	-----	431
Db	17255	RVLQGLLKPLFKSTSVGPLYSGCRLTLRLPEKHGAATGVDAICTLRDPTGPGDRERLY	17314	Db	18335	QHPGSRKFNTTERTVLOGLLTPLFKNTSVGPLYSGCRLTLRLPEKQBAATGVDTICTHRVD	18394
QY	409	-----	408	QY	432	-----	431
Db	17315	WELSQLTNSITELGPYTLDRDLSLVNGFNPWSSVPTTSTPGTSTVHLATSGTPSSLPGHT	17374	Db	18395	PIGPGLDREXLYWELSXLTXIXELGPYXLDXSLVNGFXXXXXXXTSTPGTSXVXLX	18454
QY	409	-----FORTFIDPL-----	417	QY	432	-----	431
Db	17375	TAGPLLVPTFLNFTITNLKYEEDMCPGSRKFNTTERTVLOQLHGPMPFKNTSVGPLYSGCR	17434	Db	18455	TSGTPXXXPXTXXPLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERTVLOGLLXPXFK	18514
QY	418	-----WD-----Y	420	QY	432	-----	431
Db	17435	LTLRSEKGAATGVDAICTHRLDPKSPGLDREXLYWELSXLTXIXELGPYXLDXSLY	17494	Db	18515	XTSVGXLVSGCRLTLRLXEXKXXAATVXDXXCXXDXDPXPGLDREXLYWELSXLTXIXE	18574
QY	421	L-----	421	QY	432	-----	431
Db	17495	VNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXPXTXXPLXPFTLNFTITNLXYEEXM	17554	Db	18575	LGPYXLDXSLVNGFTHRESSVPTTSSPGTSTVHLATSGTPSSLPGHTAPVPLLIPTFLN	18634
QY	422	-----	421	QY	432	-----	431
Db	17555	XXPGSRKFNTTERTVLOGLLXPXFKTSTVGXLYSGCRLTLRLXEXKXXAATVXDXXCXXDX	17614	Db	18635	FTITNLHYEENMOHPSRKFNTE	18694
QY	422	-----DSL-----	424	QY	432	-----	431
Db	17615	PXXPGLDREXLYWELSXLNSITELGPYTLDRDLSLVNGFTHRSSMPTTSTPGTSVHL	17674	Db	18695	TGVDAICTLRDPTGPGLDREXLYWELSXLTXIXELGPYXLDXSLVNGFXXXXXX	18754
QY	425	-----LF- 426	426	QY	432	-----	431
Db	17675	TSGTPASLPGHTAPGPLLVPTFLNFTITNLQVEEDMRHPSRKFNTE	17734	Db	18755	TSTPGTSXVXLXTSGTPXXXPXTXXPLXPFTLNFTITNLXYEEXMXXPGSRKFNTTE	18814
QY	427	-----LDEIR-----	431	QY	432	-----	431
Db	17735	STSVGPLYSGCRLTLRLPEKGAATGVDTICTHRLDPLNPGLDREXLYWELSXLTXIXE	17794	Db	18815	RVLQGLLXPXFKTSTVGXLYSGCRLTLRLXEXKXXAATVXDXXCXXDXDPXPGLDREXLY	18874
QY	432	-----	431	QY	432	-----	431
Db	17795	LGPYXLDXSLVNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXPXTXXPLXPFTLN	17854	Db	18875	WELSXLTXIXELGPYXLDXSLVNGFTHRTSTVPTTSTPGTSTVHLATSGTPSSLPGHT	18934
QY	432	-----	431	QY	432	-----NF-----	433
Db	17855	FTITNLXYEEXMXXPGSRKFNTTERTVLOGLLXPXFKTSTVGXLYSGCRLTLRLXEXKXXA	17914	Db	18935	APVPLLIPTFLNFTITNLQVEEDMRHPSRKFNTE	18994
QY	432	-----	431	QY	434	-----SAR-----	436
Db	17915	TXVDXXXXXDPXPGLDREXLYWELSXLTXIXELGPYXLDXSLVNGFHPRSSVPT	17974	Db	18995	LTSLRPEKXGAATGMDAVCLYHPNPKRPGLDREQLYCELSQLTHNITELGPYSLDRDLSLY	19054
QY	432	-----	431	QY	437	-----	436
Db	17975	TSTPGTSTVHLATSGTPSSLPGHTAPVPLLIPTFLNFTITNLHYEENMOHPSRKFNTE	18034	Db	19055	VNGFTHQNSVPTTSTPGTSTVYWAATGTPTSSPFGHTXXXPLXPFTLNFTITNLXYEEXM	19114
QY	432	-----	431	QY	437	-----	436
Db	18035	RVLQGLGPMFKNTSVGLYSGCRLTLRLPEKNGAATGMDAICSHRLDPKSPGLDREXLY	18094	Db	19115	XXPGSRKFNTTERTVLOGLLXPXFKTSTVGXLYSGCRLTLRLXEXKXXAATVXDXXCXXDX	19174
QY	432	-----	431	QY	437	-----	436
Db	18095	WELSXLTXIXELGPYXLDXSLVNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXPXT	18154	Db	19175	PXXPGLDREXLYWELSXLTXIXELGPYXLDXSLVNGFTHWSGLTSTTPTWTSTVDLG	19234
QY	432	-----	431	QY	437	-----SPT-----	439
Db	18155	XXXPLXPFTLNFTITNLXYEEXMXXPGSRKFNTTERTVLOGLLXPXFKTSTVGXLYSGCR	18214	Db	19235	TSGTPSPVPTTAGPLLVPTFLNFTITNLQVEEDMRHPSRKFNATERVLOGLLSPIFK	19294
QY	432	-----	431	QY	440	-----	439
Db	18215	LTLRLXEXKXXAATVXDXXCXXDXDPXPGLDREXLYWELSXLTXIXELGPYXLDXSLY	18274	Db	19295	NTSVGPLYSGCRLTLRLPEKQBAATGVDTICTHRVDPGPGLDREXLYWELSXLTXIXE	19354
QY	432	-----	431	QY	440	-----	439
				Db	19355	LGPYXLDXSLVNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXPXTXXPLXPFTLN	19414



QY	440	-----	439
Db	19415	FTITNLXYEEXXXPGSRKPHRRNTTTERVLOGLLXPFXKTSVGLYSGCRLTLRXXEKX	19474
QY	440	YN-----	442
Db	19475	XAATXVDXXCXXDXPGLDREXLYWELSLTXIXELGPYXLDXSLYVNGFTHRSF	19534
QY	443	-----	442
Db	19535	GLTTSPTWSTVDLGTSGTSPVPSPPTAGPLLVPTFLNFTITNLQYEDMHRPGSRKEN	19594
QY	443	LTP-----	445
Db	19595	TERVLOGLLTLFRNTSVSLYSGCRLTLRPEKDGAAATRVDACTHRDPKSPGLDRE	19654
QY	446	-----	445
Db	19655	XLYWELSLTXIXELGPYXLDXSLYVNGFXXXXXXXTSTPGTSKXVLXTSGTXXXXP	19714
QY	446	-----	445
Db	19715	XYTXXPLXPFTLNFTITNLXYEEXXXPGSRKFTTTERVLOGLLXPFXKTSVGLYS	19774
QY	446	-----	445
Db	19775	GCRLTLRXXEKKAAATXVDXXCXXDXPGLDREXLYWELSLTXIXELGPYXLDX	19834
QY	446	-----	445
Db	19835	SLYVNGFTHWIPVPTSTPGTSTVDLGTSGTSPSPPTAGPLLVPTFLNFTITNLQYGE	19894
QY	446	-----	445
Db	19895	DMGPGSRKFTTTERVLOGLLGPFXKTSVGLYSGCRLTLRSEKDGAAATGVDAICHH	19954
QY	446	-----	445
Db	19955	LDPKSPGLDREXLYWELSLTXIXELGPYXLDXSLYVNGFXXXXXXXTSTPGTSXVK	20014
QY	446	-----	445
Db	20015	LXTSGTXXXXXXTXXPLXPFTLNFTITNLXYEEXXXPGSRKFTTTERVLOGLLXP	20074
QY	446	-----	445
Db	20075	FXKTSVGLYSGCRLTLRXXEKKAAATXVDXXCXXDXPGLDREXLYWELSLTXIX	20134
QY	446	-----	445
Db	20135	XELGPYXLDXSLYVNGFTHQTFAPNTSTPGTSTVDLGTSGTSPSPSPTAGPLLVPT	20194
QY	446	-----	445
Db	20195	LNFTITNLQYEDMHRPGSRKFTTTERVLOGLLGPXKNTSVGLYSGCRLTLRPEKNG	20254
QY	446	-----	445
Db	20255	AATRVDACTHRDPKSPGLDREXLYWELSLTXIXELGPYXLDXSLYVNGFXXXXXX	20314
QY	446	-----	445
Db	20315	XYTSTPGTSKXVLXTSGTXXXXXXTAPVPLLPFTLNFTITNLHYEENMHRPGSRKFT	20374
QY	446	-----	445
Db	20375	TERVLOGLLRLPLFKTSVGLYSGCRLTLRPEKGAATGVDAICTLRDPTGPGLDRE	20434
QY	446	-----	445
Db	20435	LYWELSQLTNSVTELGPYTLDRDSLXVNGFTQSSVPTTSIPGTSVHLETSPTASLPG	20494
QY	446	-----	445
QY	446	-----	445
Db	20495	HTAPGPLLVPFTLNFTITNLQYEDMHRPGSRKFTTTERVLOGLLKLPLFKSTSVGLYSG	20554
QY	446	-----	451
Db	20555	CRLLTLRPEKGAATGVDTICTHRLDPLNPGLDREXLYWELSLTXIXELGPYLLDRGS	20614
QY	452	-----	451
Db	20615	LYVNGFTHRNFPVPTSTPGTSTVHLGTSETSPSPRPVPGPLLVPFTLNFTITNLQYEE	20674
QY	452	-----	451
Db	20675	AMRHPGSRKFTTTERVLOGLLRLPKNTSIGPLTYYSRCRLTLRPEKKAATRVDAICT	20734
QY	452	-----	451
Db	20735	HHPDFQSPGLNREXLYWELSQLTHGITELGPTTLDRDSLXVVDGFTHWSPIPTSTPGTSI	20794
QY	452	VNLSTS-----	457
Db	20795	VNLGTSGIPPSLPETTTXXXPLLPFTLNFTITNLXYEEXXXPGSRKFTTTERVLOGLLK	20854
QY	458	-----	457
Db	20855	PLFKSTSVGLYSGCRLTLRPEKGVATRVDAICTHRDPKIPGLDRQOLYWELSQLTH	20914
QY	458	-----	457
Db	20915	SITELGPYTLDRDSLXVNGFTQSSVPTTSPTGTFTVQPETSETPSSLPCTATGVLPLP	20974
QY	458	-----	457
Db	20975	FTLNFTITNLQYEDMHRPGSRKFTTTERVLOGLLMPFKNTSVSLYSGCRLTLRPEK	21034
QY	458	-----	457
Db	21035	DGAATRVDACTHRDPKSPGLDRERLYWKLSQLTHGITELGPTTLDRHSLYVNGFTHOS	21094
QY	458	-----	457
Db	21095	SMTTTRTPTDSTMHLATRTASLPGTTPASPLLVFTLNFTITNLRYEENMHRPGSRKP	21154
QY	458	-----	457
Db	21155	NTTTERVLOGLLRLPVFKNTSVGLYSGCRLTLRPKDGAATKVAICTYRDPKSPGLDR	21214
QY	458	-----	457
Db	21215	EQLYWELSQLTHSITELGPYTQDRDSLXVNGFTQSSVPTTSVPGTPTVDLGTSGTPVSK	21274
QY	458	-----	457
Db	21275	PGPSAASPLLVFTLNFTITNLRYEENMHRPGSRKFTTTERVLOGLLRLFKSTSVGLY	21334
QY	458	-----	457
Db	21335	SGCRLTLRPEKDGATGVDAICTHHPDKSPRLDREXLYWELSQLTHNITELGHVALDN	21394
QY	458	-----	457
Db	21395	DSLFGVNGFTHRSSVSTTSTPGTPTVVLGASKTPASIFGPSAASHLLILFTLNFTITNLRY	21454
QY	458	-----	457
Db	21455	EENMHRPGSRKFTTTERVLOGLLRLPKNTSVGLYSGSRLTLRPEKGEATGVDAICTH	21514
QY	458	-----	457
Db	21515	RPDPTGPGLDREXLYWELSQLTHSITELGPYTLDRDSLXVNGFTHRSSVPTTSVGVVSEE	21574
QY	458	-----	457

Db 21575 PFTLNTINNLRAYMADMGQGSGLKFNITDNNVMKHLSPFLQSRSSIGARYTGCRVIALRSV 21634  
QY 458 -----  
Db 21635 KNGAETRVDLLCTYLQPLSGPLPIKQVPHLSQQTHGITRLGPKSLDKDSLNGYNEP 21694  
QY 458 -----  
Db 21695 GUDPEPTPKPATTFPLPPLSEATTAMGYHLKTLTNFTISNLIQYSPDMGKGSATFNSTEG 21754  
QY 458 -----NSLW 462  
Db 21755 VLQHLRLPLFKQSMGPFYLGQLISLRPEKDGATGVDTTCTYHPDPVPGGLDIOQLYW 21814  
QY 463 -----  
Db 21815 ELSQLTHGVTQLGFYVLDSDLSIFNGYAPQNLISIRGEYQINPHVNNLSNPDPTSSEVI 21874  
QY 463 -----  
Db 21875 TLLRDIQKVTTLYKGSQSLHDTFRCLVTNLTMDSVLTVKALFSSNLDPSLVEQVFLDK 21934  
QY 463 -----WL 464  
Db 21935 TLNASFHWL 21943  
RESULT 10  
AAO29660  
ID AAO29660 standard; protein; 8601 AA.  
AC AAO29660;  
DT 03-SEP-2003 (first entry)  
DE Paederus fuscipes Pedf mixed type I PKS/NRPS.  
KW Pederin; polyketide synthase; PKS; nonribosomal peptide synthase; NRPS;  
KW antitumour agent; beetle; Pedf.  
OS Paederus fuscipes.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 7136  
FT /label= Unknown  
FT /note= "Encoded by RTC"  
XX  
PN WO2003044186-A2.  
XX  
XX 30-MAY-2003.  
XX  
XX 21-NOV-2002; 2002WO-EP013085.  
XX  
XX 22-NOV-2001; 2001EP-00127395.  
XX  
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
PA  
PI Piel J;  
XX  
XX WPI; 2003-468769/44.  
DR N-PSDB; AAL60188.  
XX  
PT Novel gene cluster encoding modular polyketide synthase enzyme involved  
PT in biosynthesis of antitumor compound pederin, useful in preparing  
PT modified pederin biosynthesis gene cluster, or a modified pederin  
PT molecule.  
XX  
PS Claim 10; Page 46-69; 90pp; English.  
XX  
XX The invention relates to nucleic acid comprising pederin biosynthetic  
CC gene cluster containing pederin units which are either coding units  
CC for individual enzymes or for one or several polyketide synthase or

CC nonribosomal peptide synthetase modules. The nucleic acid of the  
CC invention is useful in the preparation of a modified pederin biosynthesis  
CC gene cluster, or in the preparation of a modified pederin molecule. The  
CC modified pederin molecules are useful as alternative antitumour agents  
CC and are even more potent antitumour agents than the original pederin. The  
CC present sequence is Pedf mixed type I polyketide synthase/nonribosomal  
CC peptide synthase (PKS/NRPS) encoded by pederin biosynthetic gene cluster  
CC from Paederus fuscipes beetle. The polyketide comprises module 1 PKS (KS  
CC -ACP), module 2 NRPS (C-A-T), module 3 PKS (KS-KR-ACP), module 4 PKS (KS-  
CC KR-MT-ACP), module 5 PKS (KS-KR-DH-ACP), module 6 PKS (KS-KR-ACP),  
CC module 7 incomplete PKS (KS-DH), where KS is ketosynthase domain, ACP is  
CC acyl carrier protein domain, C is NRPS condensation domain, A is NRPS  
CC adenylation domain, T is NRPS thiolation domain, KR is ketoreductase  
CC domain, MT is methyltransferase and DH is dehydratase domain  
XX  
XX Sequence 8601 AA;  
Query Match 22.0%; Score 539; DB 6; Length 8601;  
Best Local Similarity 2.7%; Pred. NO. 0.0011;  
Matches 225; Conservative 82; Mismatches 133; Indels 7873; Gaps 73;  
QY 1 MGRKEMVVRDVPKMFVLIS-----ISFLVSVFINCKVM 33  
Db 235 VGANLLR--PEPFVLLSESGQLSESASVHSFGAQAQGHRAEGVCSLLI-----KPL 286  
QY 34 SKAL-----  
Db 287 TKALADGDPYASIKHSVAVNFNGQGGASIAAPNVSHVDLIKSCYQARVDPQVRYIEA 346  
QY 38 -----YNR-----PWRG----- 44  
Db 347 QGMGNVLADLVEMQAFNRALTDIARQQRVSLPPGNCILSTLTKPMGHMESASALGALPKV 406  
QY 45 -----  
Db 407 IRSLHTRTIHKIAHTQYHPDMYQGPQCATAGETVAMPQMEGLRAGTHCYMGGVNAH 466  
QY 45 -----LVLS----- 48  
Db 467 LLVESVAGYDDSELGTGVSSLLEHLVILVLSAKTSESLRMMARRLQQLQKADAVPALRD 526  
QY 49 ----- 48  
Db 527 IAYTLQVGRDAFEHRLALVDSQQQLIEGLCYLEEROPSGEGAVYOGQVASESQSLPF 586  
QY 49 ----- 48  
Db 587 TEDDLAAVARCWAGAVLWVPVGPVKPRVRRLPAYPFKRYWVDSAVVEAERAPNSK 646  
QY 49 -----KIGKY----- 53  
Db 647 APASMLSGERSIGDYLRAKLGEVLQVPVERIDPQOHLVDLGVDSIVAMKLENLARAEGI 706  
QY 54 ----- 53  
Db 707 PVRGRDLLQYSTVQALSRLHQAQLDRDQVSVGEDEEPRQLMASRRCSLSEQKGLWVLQ 766  
QY 54 ----- 53  
Db 767 QLASEMTAYNIPLCVRIAQVLDITALEAFALLQYPLITSVFVDNGELFRECHVAAA 826  
QY 54 -----KLDQ-----LKLEILR----- 64  
Db 827 LPFWOETNTLDQAQVRMLKCLAKQPFLEKEKGLVRLHLVLSGSEHDHYLLCVHHIVPD 886  
QY 65 ----- 64  
Db 887 GGSFLPVFGLLQTYQLISQGGTFLAKSTRTGEYADFLWEQRMASAEGBQRHAYWKQ 946  
QY 65 -----QLETTISTKY-NVSKO----- 79  
Db 947 LSGELPVLSLFTDNPRDAGQRTGTYGFDLVNLSRKIRNFAKQQRNLSTLFLALPKL 1006

QY	80	-----	79
Db	1007	LLHRYSGSELIGMPEQSRSERFEGVGVYFVNMLPIRSRGVSKPLAEFARDLQLSMA	1066
QY	80	-----	79
Db	1067	DAMDHAVYPPVVMVRDLGRAPAEIDLAPIQVAFEPYQNVFSAQDLRLFNQSYRESLGVTFL	1126
QY	80	-----	PV 81
Db	1127	BEFVQGEYELALEVREGEIDFALNLKFNPTLYRMATIARMAEHLILAEHAIDAPLSPC	1186
QY	82	KNLTWNT-----	92
Db	1187	RELTMUSERERHLLHEWNATTEPYSCFHOLFQKQARMPQIAIAIFQEQRLSYAELD	1246
QY	93	-----	92
Db	1247	ERSERLAIYQQGVQPNRIVAVCLERSLDMVALIGIARSGNAWLPLDPNYPDDRLRFM	1306
QY	93	-----	92
Db	1307	LSDSQAQLLTBEGLRDKTAAIVSQAVGERLQIVAMDGHWPETERQARTSELQMRDPRN	1366
QY	93	-----	92
Db	1367	LAYVIYTSGSTGIPKGVMIHRSLVNFYMLNRPGLRANDRLAVTTYCFDIAGLELIV	1426
QY	93	-----	92
Db	1427	PLLCGACCCICATDKLNDSEALQGEIERLOPTMQATPWTLLFHGWNRRQGVKILCG	1486
QY	93	-----	XYILAG 98
Db	1487	GEPLPALQRPAESASQTNLFGPTETTITWSTVSRDLTKDSVDIGTPIANTRYVILNG	1546
QY	99	-----	98
Db	1547	DDQLVPVIGVGBELCTAGDGLARGYLGNPQLTAQKFIANPEPGRNLYRTGDLARWEDGV	1606
QY	99	-----	98
Db	1607	LEHLGRDQOVKVRGRIEISDIETWLNHRPSVAQSLVVGHEQAGMMLLVAYYVDSEWA	1666
QY	99	-----	108
Db	1667	SVSSTELRNYLAEHLPEYVMPAFFRALSNNPLMNGKVDKALSARELVAESESSEGRLP	1726
QY	109	-----	108
Db	1727	CSDIEREVLDIWRSLLAVEGIGSVGFVEVGGNSILSVMLAQOISEAFGIRPAATDLFKY	1786
QY	109	-----	108
Db	1787	PTIRDISLLIGETRESSETKTGTAGDGGKADSVLQAQRGRQSVTGYDPYQDSLAI	1846
QY	109	-----	109
Db	1847	GISCNWPARTLRQFWENLRQKESSTLSRELRAGVPEELIRHPDPVPMQYSMEGKE	1906
QY	110	PD-----	FYSTOL 117
Db	1907	LPDPPDFNLAKNALFMDPQYRVLQQAQWAIEDAGYVAQDIPETAVFMSANNFYKTL	1966
QY	118	R-----	118
Db	1967	HSAGAVETDEYAAMIAGGGTIPTWISYQLGFKGSPFAVHNSCSSLVGLYLASQCLRL	2026
QY	119	KPAKY-----	123
Db	2027	KEAKYALVGGATLFPVAGTGHLYTTPDMNLSSDGHCKAFDADADGLVGGEAVVLMVRKAL	2086
QY	124	-----	123
Db	2087	DAIRGDPYIALIRGVAVNNDSKVKGFYAPSVNGQAAVIOKALDITVDPOSVAYVEAH	2146
QY	124	-----	133
Db	2147	GTGTRLGDPVEIMALNEVYRRYTEKQPCRICSVKPNIGHLDITVAGLAGMLKVVLKHA	2206
QY	134	-----	152
Db	2207	EFPPSINTREPNPAIDFTSSPPEVVTQLTTPMPAGNEPRAALSSFGIGGTTNTHAILEYV	2266
QY	153	-----	152
Db	2267	ARTDADRWNEDNOGVPLPEQVVVLSAKTORLOASVVVLYLLRAQATTEQLDQLAVT	2326
QY	153	-----	152
Db	2327	LQVGRQMDWRVAFVLVKDLHDLSEKLERFLOGDSLQDCFCQGRVATSVMDAAATPLPVAQ	2386
QY	153	-----	152
Db	2387	DREQAAIAKAWVTGRLDVWKELPRRGTPHRISLPTYPFAEERYVWEMPPELPGRSESEQT	2446
QY	153	-----	152
Db	2447	KEWIEGOAERTLLVVHPLWQAHAVVVRERPLIFTEHLVWLCGFDVSLVRALTRCLPEGYR	2506
QY	153	-----	155
Db	2507	IVSLTPBGRGVNAQRYQSLCLQMLERLQKIGDATKTLIQLVLPOGEYSLSFGLHALLK	2566
QY	156	NVSKN-----	161
Db	2567	TVSQENPKVAQLIRVSGSETARSLADKLIENTFAPDDSHLYAPSRMLDQTLRQEST	2626
QY	162	-----	161
Db	2627	VLDMPWKEGGVYLLTGGAGGLGVLPABEIRARVRKATLVLRQSPLSAAABAERIAALR	2686
QY	162	-----	169
Db	2687	SDSITVICRQADISCATSCSCLIADIAEQ-CGTIDGILHTAGVVRDAPFLNKSAAEFQBV	2745
QY	170	-----	169
Db	2746	LAQVAGTVNLDRATQALGLDFFLLFSSAAAAGNAGQADYCAANAFDAYAYERNORVA	2805
QY	170	-----	169
Db	2806	AGQCRGHTLSVGMPLWRDGMRLNBEAQOAMRYTTGLVPMDSRSGIRGLYRSLAARLGT	2865
QY	170	-----	169
Db	2866	LVLEGDATAIGSLLANGTARSVSELGVPAAANGNDLDELKOKTIYQLKLLAQVIGRAVE	2925
QY	170	-----	169
Db	2926	RIESCEPMDRYGLDSIAITQLNRKLEEQFGLSKTLFYQYQTVAEALAEVYLNLKTVSCRA	2985
QY	170	-----	169
Db	2986	WTGLRDESVLVADAARRGLPLPETAPVVERNVLPGVNAVQEPPIAIGLSGRYPOAETLEE	3045
QY	170	-----	181
Db	3046	FWENLQAKDCVSEIPEDRWLENFPHDPKPAVQAQKSYKWKGGFIEGFAEFDPLFFNI	3105
QY	182	PR-----	WN 185
Db	3106	SPREALANDPOERLFLQCAWHVLEADAGYTROSLOQGGHKVGVFVGTITKTGFDLYGPBLWH	3165
QY	186	-----	185

Db 3166 RGERLPHTSFSSVANRVSCLNLKGPSMPIDTMCSSSLTAIHEACQHLRQSGCDMAIUVG 3225  
QY 186 -TKLYYVGP- 193  
Db 3226 GVMYVHPSTYVGLCSAYMLSDGQCRSPQGGNGFVPGEGIGAVLLKPLARAQEDDLI 3285  
QY 194 - 205  
Db 3286 HAVIRSSVNHGRTNGYTPVPNPAQELIGDCLKKAGVDARSIGYIEAHGTGTGELGDI 3345  
QY 206 -GLTALL- 212  
Db 3346 EVNGLAQAFQAGEHSRCFLGSVKSNLGHLFAAGMAGLTKVILQMRHGQIVPSLHAQV 3405  
QY 213 - 212  
Db 3406 LNPNIIDFATPTVPQOLVEMWRTILQESGRSRELPRRAGLSSFGAGGSNAHLILEEYIA 3465  
QY 213 - 212  
Db 3466 PEPAORFGEPTAAVILLAKTPECLRRVVSLLAFIESELTRTPDQTLFDIAYTL 3525  
QY 213 - 212  
Db 3526 QVGREALDERLGLNAVLSQELSRQAAFLGEEAEQPLLYGRVQRNKDALQALANDEEFQ 3585  
QY 213 - 212  
Db 3586 ETVDKWLARRKYSKLLKFWVTGLSDWTRLYSVDLPRRLPYPPVRQRYWLDAYSLEP 3645  
QY 213 - 212  
Db 3646 MYPTEQSPVVPVDAEVSQDAGREADMLMGPVDAVVBQGTEDFPFAGARIAMVGGSEA 3705  
QY 213 - 212  
Db 3706 QKRAVFEQPKALELAAGAVGAASIALRGLDHHVWFAPASQTQGWADERIIDAQRDGVL 3765  
QY 213 - 212  
Db 3766 ALFQLVKLLABGYGVAEFGMTVITQALATCDTERIDPTHAHVGLVGLAKEYPKWRL 3825  
QY 213 - 212  
Db 3826 RALDIDARAEVWPGLWRLPHTRGESVYRWGCEWLRLQRLVALNGMPVAKGRAYRRQGY 3885  
QY 213 - 212  
Db 3886 VVIGGAGLGMTWSRMIRDHQAQIVWLGRSAKDATVRAKLDEVADDGLAPDYWQIDARD 3945  
QY 213 -RYAQ- 216  
Db 3946 ADALRQTFQVREYQIHGVSTLGDYDQSVQWSEALFREILLSSKLDIGVRLSQCLR 4005  
QY 217 - 216  
Db 4006 DEALDFVFFSWVAFRSGGMAAYSACAFTDFAQLGNELACAVKVINWYNNLGGG 4065  
QY 217 -RNC- 219  
Db 4066 TRISAALKRLVQRGVRPIEAREGLCALAVLLDGLRLQLAVTRTCOPAAIETFEAGQWL 4125  
QY 220 -THSFY- 234  
Db 4126 VKAGTHSCFANVEAYOPTQMPQESPDAARLNIWIRLLFVQLQSILGFQETGFQFNATAI 4185  
QY 235 - 234  
Db 4186 RRQAGIVDKYRWRESNLIAEHGYRLAGDEVARIASADEIGESSRRLWQEWRECKT 4245  
QY 235 -RVPKYINGTKL- 245  
Db 4246 RFLEQPOHTLAVLVEDCLSQLPEVLRGTRLVTDILFPNGSMEKIEGLYKNNLICDYFND 4305

QY 246 - 245  
Db 4306 VVAGVAQAYIORRENEPNAEIRLLEVGAGTGTSTVLPQLNLWRAFIAYTDLSSK 4365  
QY 246 -KNTWRKLK- 253  
Db 4366 FPNHARLYGTDYPYITYRLLNIEBPLIQDDIEITYDILIATNVLHATNRMENTURNAK 4425  
QY 254 - 253  
Db 4426 AALRNGIILLNEISDKTIPASVLPGLIDGWSLAEDHWRIPGSPQLPAENWQALLQSG 4485  
QY 254 -R- 254  
Db 4486 FDKVSFPAQVAHDLGQIIVAQTNQVHIHQHAGPVLETAADKPLPTLESAAERLVD 4545  
QY 255 KOAPV-KEQFEKKAKTQS- 272  
Db 4546 SSVPARRODVAARVELILDSLAQALSTGREQIE-QDIPFSDYDIDILGVGVQ 4599  
QY 273 -TTPYFSYTT- 282  
Db 4600 RLNDELGLSLNTLILFDYTTVQRLAEHIVAEYGHITLDVPAALPGPELSVSEPAIDIP 4659  
QY 283 - 282  
Db 4660 VQAVPSLPRREAVVQTDGIAVIGMAGOPPGADSVDALWQNMVAGVNPVTELSLYL 4719  
QY 283 - 282  
Db 4720 AVSPEKQPKSYCKWGGALQGRDCPDPLFFNISPREASMNPHQLIQESWKALEDAGY 4779  
QY 283 - 282  
Db 4780 APRSLDSRTGIFVCAEPSAYVHESFVGASDAIVASRLSYFLDLKGPVAVNTGSSGV 4839  
QY 283 - 282  
Db 4840 ALHLACESLRNGETEVALAGGVAVMGOTILVGLAQTDMLSRGTCCTPDADAGWVSE 4899  
QY 283 - 282  
Db 4900 GVMVVLKRLDQALSDDTIYGVIRASGINDGASNGITAPSGIAQQQLITVYRYAID 4959  
QY 283 - 282  
Db 4960 PRITYVEAHGTGTGLDGPVEANALVRAFRSFTSTGYCAVGSIKSHIGHTSSSGVIGL 5019  
QY 283 -SAALN- 287  
Db 5020 ISILLCLKHHQLPGMRHFRLNPLIEFERSPFYVNMWMPWRSGEPLMAALNSFGHSG 5079  
QY 288 -VTN- 291  
Db 5080 TNVHLVVEEFVRSNSEDPRVDDVSSTAQPELILLSTKDAERLSEVLNLAHFVQAQO 5139  
QY 292 -VTYSITTA- 299  
Db 5140 PADLERLSLADLAYTLQGREAMEQVALLVCDLAGLLEALSALREERPCVSVWGRVE 5199  
QY 300 -ARRV- 303  
Db 5200 PGPSRGAETVNADQPAELLORIPOWLAEGALDELAQWAGAPIDWCQLRRRPPRRVH 5259  
QY 304 -STST- 307  
Db 5260 LPSYPFARERYWRSEPAVHSPVVAAGLHPLVORNTSTLDRHCFESSFGSEFFFRDHRVQ 5319  
QY 308 - 307  
Db 5320 QGPLLPAYLEWARAAQIALGNACPDVALKLSNVWVIGPLLAEQPIVGTITLQAREDR 5379

QY 308 ----- 307  
Db 5380 GIDYQISSVSAAGQPVVHCQIATTETETEKEAAPVLDDLALRSRLTQKEIGVERCYAALE 5439  
QY 308 ----- 307  
Db 5440 AAGVNHGPMRGLQAVSRNAEEVLATLRLPAETVGEASAVLHPALDAALQASIALTLR 5499  
QY 308 ----- 307  
Db 5500 DDEVPSPETSPRPVLLPFALESRLVVPCCASMAWIRLVAVEHAGQALQRLDVIDCTK 5559  
QY 308 ----- 307  
Db 5560 EGEVCVALRGFTSRSLPPSGATESRASASAASSTLVSTEGVSRFGKEEFLRDHSQMLPA 5619  
QY 308 ----- 307  
Db 5620 AVYLEMVRAPAEKGHERKITGLSHVWPVKVLLVSGEGREVRTCLTNVDRSAFLISACEQS 5679  
QY 308 ----- IAYR ----- 311  
Db 5680 SEGPOEVTYCOGNLLPEVMEEPGAALTAIBAIAYRCPVLEAKQCDRLQLQSTHGPALMSV 5739  
QY 312 ----- 311  
Db 5740 QQLRYSDREALLOLPDELQMGDDYGMHPSLLNGAILASVVMCLARAPRSRAGLPMPF 5799  
QY 312 ----- 311  
Db 5800 SLDLRVFQPFERQMAQYVRRHGSARSNGENLEKVDIDLDSQGRCLASLEGTLVFPD 5859  
QY 312 ----- PD 313  
Db 5860 ANRLVVAIPQWVEQALPARVAASAPLAVQAPVILAGAGEPLRRALHDSWPDALLHELPE 5919  
QY 314 SSP ----- 316  
Db 5920 SAFEVGDLRQAVVEFGMCRLLPYKGAALQPLLVLLPEAREVTPQALLGALSGLLK 5979  
QY 317 ----- MK ----- 318  
Db 5980 TVRLEHPRITARIISYPVDDTVTAGMWKVLAAEIASPEGDVEIRYDQARRHKVLHEIT 6039  
QY 319 ----- 318  
Db 6040 LSAGEHGDLSLPRDDVVMLTGLGGIGRQIARVLGVERRVRLALSRSALDDKGERFLQE 6099  
QY 319 ----- 318  
Db 6100 LRREGAVSVLYRVDVADADAVGRALLAIEQEHGGLTGIIHSAGIIADDVLSKTTAQFEQ 6159  
QY 319 ----- 318  
Db 6160 VLKPKVGVVNLDAATANRSLRYLLVFSSTAGVLGNMGQADYAAANGFLDSFAHYREALV 6219  
QY 319 ----- 318  
Db 6220 RQGLRSKLSLNLWPLWREGGMQRHGEALMQOATGMLAMESAQFEALEAGLSAQAO 6279  
QY 319 ----- 318  
Db 6280 ILVAFGEVPSIRNRLTFRMDAPEPPAPSVEMDRAPGEVGEQETQQLVRSVEBELIRI 6339  
QY 319 ----- SIMAT ----- QLRDLA --- 329  
Db 6340 VAFVQIPAEKINVRDISAYGDSISPTFEFANALNKAYKLSIMPTLFFEFIASLADLAGH 6399  
QY 330 ----- 329  
Db 6400 LLTQHRPALLEKHAVDVEKPECHSGVAAQIPIPTWPKSECIATLPLLPVGSIEPEPQAD 6459  
QY 330 ----- 329

Db 6460 LEAVAVIMAGKFPGCCDLEDFTWCQSCODLISEVPEQRMDWRRFYGDPHQBPGRKTK 6519  
QY 330 ----- TW ----- 331  
Db 6520 WGGFVADACFDARFFGISPVEAEVMDPOLRLPLETVMAALEDAGYPAGLAGSRTGVFA 6579  
QY 332 ----- 331  
Db 6580 GVATADYKOLLIEARARGLVQTPSEPFPFMIANRISYWFNFNGPSEBIDTACSSSLIAVH 6639  
QY 332 ----- 331  
Db 6640 KAIESRLGSCWALAGGVNLGSPRITTIASSOAGMLSEBGRMTFDERANGYVRSEGVA 6699  
QY 332 ----- VYT ----- T 335  
Db 6700 ILLKPLRKAIAADNDRIHGLIRSGENHGRSASTAPNGNAQKRLLVDIYSRADIDPRT 6759  
QY 336 LRY ----- 338  
Db 6760 ISYIEAHGTGTVLGDDPEVNGLKAAPQELYQSRGLDVPEQPHCGLSNVKANVGHLEAAAG 6819  
QY 339 ----- RQNPFC ----- PSNRT 351  
Db 6820 AVGIKVLMLKHKRIPGNPHLRPNPYLQLEGTPFYLRETLDMPQPTDVRGNPLARRA 6879  
QY 352 AVSEP ----- MKNTHVLIR ----- NETPV ----- 370  
Db 6880 GVSSFGVGSNAHVILEEYQBPQGWGSEPAYPALIVLSAKDEERLVCVAQRLLRPIRD 6939  
QY 371 ----- 370  
Db 6940 YGSELVHLHDIAVTQVGREAMPRLALAVTSLAQLADRLQTLWEQPTQTEGVQOGLVTOE 6999  
QY 371 ----- TIYG ----- TLDMSLLYNET ----- 386  
Db 7000 ABEQFDTVLGDDEDAARAAVERWVEKQYKELDAWTRGWAIDWNVLYCTDTRPRRIGLPTY 7059  
QY 387 ----- 386  
Db 7060 PFARRRYVASVPQAEADRGNSTLSEBPEQRSKSDLLTFEEYWAEPVLAAPATDRVKTL 7119  
QY 387 ----- MFVENKTASDNKTITPTSPSMGFQRTFIDPL ----- 417  
Db 7120 LCLCSDPEHORRIAEQXDSRDPGVOLIPIEQGDA ----- PAEPDEARQR ----- IDPLQPS 7170  
QY 418 ----- 417  
Db 7171 SYSRALTTTIKALGRVDALLYLWPCEDRRWISNVLPVLHLLQALYETGLRPRKLLSGEY 7230  
QY 418 ----- WDLDSL ----- LFLDE 429  
Db 7231 ADALERCHLDSWAFERSLGVMPTQVAVVFRERAADTGESSPTWDLVLEVLFAELFAEK 7290  
QY 430 IRN ----- 432  
Db 7291 LRSACYRQVRHVPLIRPLAWQGSAPFKQGVVLTGGGGGLGMIVAEHLATVYAARL 7350  
QY 433 ----- 432  
Db 7351 VLSGRSSSLAEBKVELLQARGAQLVYGADVTDVHAMQEWVDQARRHPSPLNGVLHIAGL 7410  
QY 433 ----- 432  
Db 7411 NGTAEVLKABADAFORVLDKATGTSQVLDQVLRRESLDFICYFCSSAIIIGDFGSDYAL 7470  
QY 433 ----- 432  
Db 7471 GNRFSQALYRAQWVSSALSCKTLINWPLWQDGLGVGDAEQTRFYLOSGGQSLCS 7530  
QY 433 ----- 432

Db 7531 QEALALLEQLTQDRAQLVWAGQPDRLRLRWVNBPLEAATVTVPVRAAKAVARAEL 7590  
Qy 433 -----PSL-- 435  
Db 7591 GGGDLQOCLLDLTKTKICELIGTQVNELEHANLVDFGDFSISLAESRVLRSFYSLDI 7650  
Qy 436 ----- 435  
Db 7651 SPSVPFSSHSTANLTAYFLAEHRQTLEGFYQQPQAGPEHAPVPTIEVAQSVVPVPTALL 7710  
Qy 436 ----- 435  
Db 7711 PGTSGSASQGBEIAIIGLSRPPQARTIEELWRILEOGRDAIQEVPIDRFWRSY 7770  
Qy 436 ----- 435  
Db 7771 SPSQEMSKSNKGGCIPGIAEFDFLFFBISLEAERMDPRQHLMOEAWLALEDAGYGP 7830  
Qy 436 ----- 435  
Db 7831 EQLECNKISMFGVEBGCDYQRRLTQQTSLTSMHNGILASRLAYFLNLKGPVMINTACS 7890  
Qy 436 ----- 435  
Db 7891 SALVAVHQACSLRHGECDTAAGVNLVLAPEAYVGMTQAGMLSPDGKCYVFDKRAEGL 7950  
Qy 436 ----- 435  
Db 7951 VPGEAVVVLKRLSKALADGDIKALIRGSGINVDKGTNGITAPSGASQTELLEGIYRQ 8010  
Qy 436 ----- 435  
Db 8011 CALQPQDISIVTHGTGLGDPIEINALVDYVFKGTDKQGFALTSIKSNLGHTPAAGS 8070  
Qy 436 -----RSPTYVN----- 442  
Db 8071 LVSLISLVLAIRHRTIPSSLHCEKNQDIYRWQESPFYVNTKXKHEWECALGPRIQAVSAP 8130  
Qy 443 -----LTPPEH----- 448  
Db 8131 GMSGTNAHVVVQEHQPAEBSRWSSTAAPYLVWVSAKTETTLQEQIRQWEDYLSRHPDLDF 8190  
Qy 449 -----RRAV----- 452  
Db 8191 EAVSYTLKGRHHFKYRCAIVAKDLSQLQALRQALDRQTQANLCMGCVDRDFSGQKAIR 8250  
Qy 453 ----- 452  
Db 8251 DFTASLAQGEALURDKPDDYRDNLIALADFYCGYEVAHVHLFAGRPQRSLPGYPPARE 8310  
Qy 453 -----NLST----- 456  
Db 8311 HYWIDFSAHRAELSERSDTQNLPLQNLSTLSEQRVYSAFAKGFERMLRIAHQBEL 8370  
Qy 457 ----- 456  
Db 8371 LIPTLFLEMARLAQQSLMDPVRLKNMVMACPLYQQGSDYELFSLHEKSDLLYTV 8430  
Qy 457 ----- 456  
Db 8431 EMQGEVPVCGHFGIEIDSTEALQLPVVPDIARLARSLFVVPDVALGAPENGVARISDVQA 8490  
Qy 457 -----SNSLW-----WWL 464  
Db 8491 DATTLATLNVADGEDRSWVHFPLLINAGWLL 8523  
|||

RESULT 11  
ID ABP76678  
XX ABP76678 standard; protein; 19938 AA.  
AC ABP76678;  
XX

DT 26-FEB-2003 (first entry)  
XX Streptomyces viridochromogenes Avi gene cluster polypeptide frame 2.  
DE  
XX Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
XX  
OS Streptomyces viridochromogenes.  
XX  
PN WO200268436-A1.  
XX  
PD 06-SEP-2002.  
XX  
PF 24-AUG-2001; 2001WO-EP009815.  
XX  
PR 25-FEB-2001; 2001DE-01009166.  
XX  
PA (COMB-) COMBINATURE BIOPHARM AG.  
XX  
PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
XX  
DR WPI; 2003-018650/01.  
DR N-PSDB; ABZ37515.  
XX  
PT New avilamycin derivatives, useful for treatment of infections, and  
PT nucleic acid encoding avilamycin synthesis enzymes.  
XX  
PS Example 1; Page 68-301; 319pp; German.  
XX  
CC The invention relates to avilamycin derivatives (I) with antibacterial,  
CC virucide, protozoacide and fungicide activity. (I) are useful for  
CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
CC or veterinary medicine, particularly where caused by Staphylococcus  
CC aureus. (I) are more hydrophilic than known avilamycins. The present  
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-  
CC ABZ37516)  
XX  
SQ Sequence 19938 AA;  
Query Match 22.0%; Score 538; DB 6; Length 19938;  
Best Local Similarity 1.1%; Pred. No. 0.0098;  
Matches 223; Conservative 92; Mismatches 141; Indels 19088; Gaps 71;  
Qy 2 GRKEMVRD----- 10  
|||  
Db 359 GRVERAGRDGWAGLEDRAGQPSRPGPLPPRAKPPSPRAAPLPGMSPLPPRCIPSSRPA 418  
Qy 11 ----- 10  
Db 419 PPPPPRATPRPRAACDNGVACRASPAGVLCGSTGPTXVGLCAESPIRHAVGRTL 478  
Qy 11 ----- 10  
Db 479 AMCSIRHRGRVHVADWNWIGVCVCICTEDRRRTGFPQSPQXSPLTYGKSQVVAHDRNR 538  
Qy 11 ----- 10  
Db 539 SAECRQVDPVQRDQERRAGGQLPVHRDRAQRRCRPRRTPHQVGRDLLPAHPFGDRR 598  
Qy 11 ----- 10  
Db 599 LRRHRHRAARLURGGGPGQQVPREHPXVRRDLFGHPRLQGRRRRRCRROGLAQGRHRDDQ 658  
Qy 11 ----- 10  
Db 659 HEADPFGPDDREGPAPPEGLADQEGPRAEGEGRGGQDPGEGRHPLRPRHRPCHRAQ 718  
Qy 11 ----- 10  
Db 719 RAPAPAPAHQALPLRLQRRGRADRRLQERAAARPGRPRRHLPQRQAGGRPRARRR 778  
Qy 11 -----VPM----- 14

Db 779 RGPRTPVRRPGRARPHRPRRLQHRPADLPDGRPEGIPRLDHOEGRRHRPRGRRRHPH 838  
QY 15 -----:|:|----- 14  
Db 839 RLPEGLHOGGHLHHPGGDLRLRRRPRBEGBAYGGGLCDAGRGRVPLQRVGAIPPR 898  
QY 15 -----:|:|----- 14  
Db 899 RCFQTRRSEGVDSRGRPLAVSRAGMVLGKLIHQSLPCRMRSRVYVYXVPLEGRGRIMS 958  
QY 15 -----FVLSI-----SFL----- 24  
Db 959 INRERSKALFPLIKVNDNHEAIEIVSKGNALVSAEDYAALREGSYLLRS PANARL 1018  
QY 25 -----:|:|----- 24  
Db 1019 LKAYENALAHVNVSERELIDPDSADAGSNAAGLSSRIIRAGMTTRPGSRTTARCSFAS 1078  
QY 25 -----:|:|----- 24  
Db 1079 SSRTSGATPSRGSANPSRXTTCRGRGGSTTNTASTWLRTRRSXSLPGTTDCPGG 1138  
QY 25 -----:|:|----- 24  
Db 1139 PPVLGWCWGPLEPDVSAGSVVNVPLQRLGADPPRLVBCAGQGPTRSRAPGSFPC 1198  
QY 25 -----VSPINCKVMSKALYNRP-----WRGLVLSKIGKYK----- 54  
Db 1199 LRQTSQSVITCRSGSRALSRAPARTSWG---SSGOPYRXGSRRRRCRSGPGXSRR 1255  
QY 55 -----:|:|----- 54  
Db 1256 RSRTRWPGARRCVRGTRWGLRRSARRXGVSAATRSRGGPTRHGXSWLRRRXDGXS 1315  
QY 55 -----:|:|----- 54  
Db 1316 VRRGATRAAGDRDFSGRSGAGSPSCSRASRSRSPRSAGTASATSPGTDYGRSR 1375  
QY 55 -----:|:|----- 54  
Db 1376 WBRRCRRRGLRRLPAGTRCGLRPGVPRARSSRVPSPRCRCRCHTCRRPPARTGSGG 1435  
QY 55 -----:|:|----- 54  
Db 1436 SLGRGRPGRXXSRXSGRAPSSVRAPRFPDAGVVSVOQDQGGAAWVCYLFEDLDGGQP 1495  
QY 55 -----LDOLK-----LEIL----- 63  
Db 1496 ABAAGQRPADDRKDCVTTDQDLRAGGVVVAVELLDEGLDVGGRGAPGGHDARGGE 1555  
QY 64 -----:|:|----- 63  
Db 1556 PPAERRAVVGPGRGAVLGCVPFEGGACQVDGAPGGVERLVLGRRADVPPTVLGAGRIDPVR 1615  
QY 64 -----:|:|----- 63  
Db 1616 RHGSPLAGLKAGRAVILXVGAEDLLRMARSGERTPAARRITSRNARFQOQALLGNR 1675  
QY 64 -----ROLETTISTK----- 73  
Db 1676 NKRTRAGEFLVMGVRPISLAVEHGWPTVLLYDQRELSKWARELLRTVTEQIAMAPDL 1735  
QY 74 -----:|:|----- 73  
Db 1736 LMELGKNEAPPEVAVVEMPADDLRIPIVREDFLGVLFDRPTSPGNIGSIIRSDALGA 1795  
QY 74 -----:|:|----- 73  
Db 1796 HGLIVAGHAADVYDPKSVRSSTGSLFSLPAVRVPSPEVMDVWEARRAAGTPIVLVGTDE 1855  
QY 74 -----YNS----- 77

Db 1856 HGDCDVFDDFTQPTLLLLIGNETAGLSNAWRTLCDYTVSIPMAGSASSILNAANAATALLY 1915  
QY 78 -----:|:|----- 77  
Db 1916 EAVRQIRISGTATTXPAAGGSCWTWPAPGPGSAPRSAPRAARSRLPDGPAWQPGSPASEPR 1975  
QY 78 -----KQPVKN----- 83  
Db 1976 RHPRTRTXIPLRAAAPAGSPTRAFRAGXPXTAPCRPRAGXPAPAXXRPVPSRXHRVTA 2035  
QY 84 -----:|:|----- 83  
Db 2036 PAAGRRRPSRSGSGDVLEAGAGLVPFPADPNPPOPGLSGGLHDLDERLPPPPVAA 2095  
QY 84 -----LWNTTEFP---QYYI----- 95  
Db 2096 DVLQVLVPLRTELPMWYQLGHSETGEBLPGVQGARQLQLODEVAERGVEPSILDVRG 2155  
QY 96 -----:|:|----- 95  
Db 2156 ADGAVVGGDTAQAVAVNRNLSGAGPWETSRPVSARPTDTAGPRTXTQDVFPSARPAPPE 2215  
QY 96 -----:|:|----- 95  
Db 2216 DAAGSARPSAKARRYPGPHSLHPPPHRRQRSGXIPANEDSRRLAPGLPFRPPKDTTP 2275  
QY 96 -----LAGPI----- 100  
Db 2276 VATTGAVSLGKRNWNTGENTGDKIQPGPLLRGLPPARRLPGVPRPGCRSSAAPG 2335  
QY 101 -----:|:|----- 100  
Db 2336 PLREPPGSGYCGRRSRPHARAAPTRASAAGPATGCAGTCPSQCVRSTAGRPARHART 2395  
QY 101 -----:|:|----- 100  
Db 2396 RRGRRTAHADRRRRCGWRTAAHRPPPGAPAGGRPARPPDRRACRNARSAAHVATARAR 2455  
QY 101 -----:|:|----- 100  
Db 2456 SARPARAGDSRPRAPAGRGPHRRSPAAASAGAAAPASGGNRCRSTGTASPDRRRG 2515  
QY 101 -----LAGPI----- 100  
Db 2516 GRGPAGRSRSGRGRGNPARSGARHRRDRNCRSAPRRGTHSPVHSVPLNRLRRSP 2575  
QY 101 -----:|:|----- 100  
Db 2576 EAIIVPRPRPGKLPAAAGHTGHRRLRGCRSSQVKGRRGRHRLNXLGFPECAPTYDCRTP 2635  
QY 101 -----:|:|----- 100  
Db 2636 MPXLRGRAGHVHPGVLSRNTSAGDRRGGLHRLARGRLPGGRGRGRAGRPDHRXPSTS 2695  
QY 101 -----:|:|----- 100  
Db 2696 GPTCGHPSRRDRCRPPGRGAVGAPRHLPSRGPDRRPFGLGHTGRRBAGQRHQRDRA 2755  
QY 101 -----:|:|----- 100  
Db 2756 GSGARRRGAGRVRLHRRGALRGRSGPDRGHAAQGGALRHGQVLURGEVHRPVQPAARN 2815  
QY 101 -----:|:|----- 100  
Db 2816 RAQRVAARQVRPAEPGPGXGGRHRLLAGLRGSAHGVRRRFPADPLXLRVRRRRRGVR 2875  
QY 101 -----:|:|----- 100  
Db 2876 RAVRHRPGVWNIGHQGEHAGGPPHRRRLRARPPRAPPAGRDPAQHGRHSCRR 2935  
QY 101 -----QNYSS----- 104  
Db 2936 RSGLDVRVAGERHRRHRLGLGPFRFARPAAGMTNLRSSRPWPTPIAVAAPYQNSSGV 2995

QY 105 ----- 104  
Db 2996 QRDGAVCRVSSLRSPAADWCRRASERTLPVGPQAEVAGEPAGVDIVDDRVLGQQAAR 3055  
QY 105 ----- 104  
Db 3056 AHPPSSAEXRCPHRTAGXRCGSGFGREXPGGARSRPPCRRPRSCXPRGASRR 3115  
QY 105 ----- 104  
Db 3116 GGTAPRTGTAAAGALAAVAHAHRPRAPRRGTGRRRRPPRRPPAARRRPTABAGRPIPA 3175  
QY 105 ----- 104  
Db 3176 APCVGTPTALAWPGSASSPRPRS PGWRPSRQVSPAGPAARIRHRSTGRGCAADAGGR 3235  
QY 105 ----- 104  
Db 3236 SPDGRHRPPPLDAARGCGSQSLSCPQLQYGCVPAGALASVSRHSSVELLSKKCAKE 3295  
QY 105 ----- 104  
Db 3296 SAAPTALALLVSKRRPTRTTGQVRDVGSRAPHSGEXPAGADLSRRRLPPGGRWNLSALANT 3355  
QY 105 ----- 104  
Db 3356 FSGTAPVLSRGRRTRDADGGAGHPDRRRGARI PVRGPSCPRDAGPADGARRETSGR 3415  
QY 105 ----- 104  
Db 3416 QAGRPPPGDRPGGPHVPGRTAGRQPPCGAARPEPVRPARRGVTGRRRGPHYR 3475  
QY 105 ----- 104  
Db 3476 APLGPLRLVAGORNAVHPELDRDRHRTRAGDPLAGRGCLLPALRDPAPRPAR 3535  
QY 105 ----- 104  
Db 3536 SDHVHGRHEHLWRACPEPVHLPAEPGRAGPDHPGATPLDPAGHRVPPLLALARPAGEREL 3595  
QY 105 ----- 104  
Db 3596 PVLDDGGFADPXGGRSAARRRERFLRPRPEPRPNRPVRRAAVHRSAAALQPGARP 3655  
QY 105 ----- 104  
Db 3656 AGRPPPAEARGAGAGPAGAVITPGLSPQGTADLGTGRIGWNVGSLCIADVSGKPVGEGNQ 3715  
QY 105 ----- 104  
Db 3716 SGLKTGDPPELRQYLGHMSLDGALGDPEQRGDLTVGVAVAQOEEDLELAVGEXMYLAEPF 3775  
QY 105 ----- 104  
Db 3776 QPCELVLKESIVARITGIAAGIWOAFVXAGCPMPXEVIRRIEYQAGRHHXTMT 3835  
QY 105 ----- 108  
Db 3836 XGRYHPCDEWVLDITHLDHLIASTAGLRLGAACCREQQPPRRGGXGPGCPVCARQSR 3895  
QY 109 ----- 108  
Db 3896 RKERGYSAWRHAQAPVMHLEESGIPAAGLPXNGARHLWSAGHARPVINGSLRDXKP 3955  
QY 109 ----- 108  
Db 3956 VSNFRIGDMRVLSSATKAPATARDSGSIDAFQNPGGPHLGSQDSQRVGLDEGAVG 4015  
QY 109 ----- 108  
Db 4016 HPEALHHLADVLDPLAAEVELAGDAGGALTTCGDAFDLGVQAGXGRGGSHLVLPGRV 4075

QY 109 ----- 108  
Db 4076 LRGVESAGAGPGTGGXLLCTQHSFTSSSGLSFRCSRFLDIGLFEQMGVRSAGCGLTGL 4135  
QY 109 ----- 108  
Db 4136 HDHGGPQEAATARXLRRSGVXRPHPALWRARSGRETAGNRHLAGRKTRCTSRGVVGRDP 4195  
QY 109 ----- 108  
Db 4196 NHGRRRQDXKFPYTCPSYPRRLRICGVERVETAXSMCCSRPVWTJRCGTAAEFSPGSP 4255  
QY 109 ----- 108  
Db 4256 RSPSNISICTCTRERCAACIIITRLSSSTCCSSRARACXSPTTPKTPTAKNSSTSRAG 4315  
QY 109 ----- 110  
Db 4316 SAPGRRPGSCTPSPRXPXRSSPCSPOTNAIRWFRRSSRCRTRWSAREGRRRRGVR 4375  
QY 111 ----- 110  
Db 4376 RHGRHPGTGFRGHSRDACQLRGTPXRPLRRAGQGLPLPALLGPAAPRRRPRDRQDP 4435  
QY 111 ----- 110  
Db 4436 GAPSRLGVPLRPDOHHLRAHPAGHPLRPEPGRGRGALRGPSGRPPGPHVRQQHQGRSH 4495  
QY 111 ----- 110  
Db 4496 RHPRRAARLARRVAPELRAGRVVRRGLPPRRRGAVGGGADGQPGAGDQGRGFAPR 4555  
QY 111 ----- 110  
Db 4616 LPVLDLGPQALTVGFPRTGEDVPYVPLELVRCSGGCELQVLRHTADFGLMYGEGYGR 4675  
QY 111 ----- 110  
Db 4676 SSLNRSMADHLRGKVAATGLVDLPGDLVVDIGSNDGTLAAYPADGPRLVGVDPAAATV 4735  
QY 111 ----- 110  
Db 4736 FAASYPPGVVELIPDFFAVDLLGGRRAKVVTSTAMFYDLPRPMEFMRVGRLLTDDGIWVT 4795  
QY 111 ----- 113  
Db 4796 EQSYLPAMLHACAYDVWCHEHLDYGLRQIEWMAERTGLKVVDABELTPVYGGSLSLVLAR 4855  
QY 114 ---STQLRKPA--- 121  
Db 4856 RGSSRQVNEPALARITAGETDLPYAEFARTEESRDLLEFLTASRDKGLHTLYGASTK 4915  
QY 122 ---KVVV--- 125  
Db 4916 GNVILQYCGLDETLLPCIAEVNEDKFGCYTPGTNIPIVSEBEARALEPDDQFLVLPWITRD 4975  
QY 126 ----- 125  
Db 4976 AMVARERDFLAGSGSLVFPPLPLEVVXSGGGDIARHHPRACRVPAVGAAGRAAVALA 5035  
QY 126 ----- 125  
Db 5036 RHGGGRVATRLRECHAGCAACRGVAAPRRPGGRRAGLSRAGPALGPATRGORD 5095  
QY 126 ----- 125  
Db 5096 DLORSRPFGARPRRSRGRPLRRPPLPRSRPDRAGAVHVPDRHGVXHPXRRGP 5155  
QY 126 ----- 125



Db	5156	HRGRGRNRRRLRVPRCLSGIRAVHRLPGRRRLPAVHPLGHRHERDLPRAGAADG	5215	Db	6236	RSQBPCAVGSVKSNIGHLEGAAGVAGVIKAILALDEDRIPASLLDGDPNPEIDWAGLDI	6295
QY	126	-----	125	QY	176	-----	175
Db	5216	TRDAPVPGGHLRGGRHRSRLAAAGSRPVPARRPAPGRAAHPRAHRLRRGGAR	5275	Db	6296	RLATRALPWPBRPHPRRAAVSGYGGTVAHVVLBQAPTAPAPAPAPAGTLPVPSAASP	6355
QY	126	-----	125	QY	176	-----	175
Db	5276	PPRLATPGACAGAPARRRAHTDHPRPVAVRGHRAHRHPSHLRRLPVRRHLPEDRAGP	5335	Db	6356	EALDRAAALAEVBEAGADLASVGHITLAHROSPLVHRAAVAVATGRDELAAGLALATQBP	6415
QY	126	-----	125	QY	176	-----	182
Db	5336	GGPRQTAGTRHRTVAVALGRPEDRPAATPGRRRPVGVRRRPAGGRRRTGGQRG	5395	Db	6416	APGLVTGAALPDAGRPPVWVFSCHGSQWAGMGRELLAEFPVFAEVIDELEPVFKBEIGFSP	6475
QY	126	-----	125	QY	183	R-----	187
Db	5396	GSADVGGARRRPAHRAHELPRDTGDPHHRARTRRDTRVVHHAGHRHPVRHRDRA	5455	Db	6476	RQMLLEGDHTEVDGAQTMIFAMQLGALALWRSRGVEPAAVIGHSVGEIAAAVTAGALTVT	6535
QY	126	-----	135	QY	188	-----	187
Db	5456	AARPAPRRGAHRAHPYKSEFRSLTTVRRILLEGKCRSGSRXFPQBYRONADYDQKNAD	5515	Db	6536	DGARLICRRSLLLREAGRGAMVSLPFDAAERLAGNDVAVAAIASSTTSTVISGDPG	6595
QY	136	-----	135	QY	188	-----	187
Db	5516	PCAPGFVSLKTRWVGADSRPVHRCRRLRTAVCERRMGRTRPLSGRRGRGTRARRGRH	5575	Db	6596	EVEKVGRWTDGLVVRVAVSDVAFHSPHMDPLDLRLRAADELSPSAPHTPLYTTALAD	6655
QY	136	-----	140	QY	188	-----	187
Db	5576	SRRPSPGNLAGGAEVGIADGIRSAVCEHLAPGCPGLAAPTIPAAAPGGRGHA	5635	Db	6656	PRATVTADGAYWAANLNPVRLAAAITAAEDGHRAFVELSPHPVVAHSIHETLAERGVE	6715
QY	141	-----	140	QY	188	-----	193
Db	5636	RPGDQAGLQRRVQAGTRRLPAGRSEPHQADRRRRQLRNAADRPVADGPRLHRRRCAP	5695	Db	6716	DVFGFTLERNQPEARTFRAAVGAAHCHGVSDWSALQNGNLEVLPPVPMOHRPLWRSI	6775
QY	141	-----	147	QY	194	-----	203
Db	5696	ALVLQAARLRLPALLGLOGAAGDRVAAPRXRAVPAEHHPGCHRLQRPYRPAVRHPQR	5755	Db	6776	AGARAERGHVDVDSHTLLGTGPGVAGSDLMLWHSTLDDDSRPYPGSHALNGVEIVPAAVL	6835
QY	148	MTC-----	150	QY	204	-----	203
Db	5756	LRCDRDGPYDGRGRARPGRGHRDGGRRPRLHLSHEVVEQCXMAANGVPLSRSTWE	5815	Db	6836	AVTFLAAGEBERRALQDWTMTHPVLTAGORQIQVREGEVVRVLRASRTVADAADNPAAW	6895
QY	151	-----	155	QY	204	-----	203
Db	5816	FGREIGHLRGQPSAVHGHTRDXARPPRRPRPATRHGPGGVQRRPSGSRAGLGRMM	5875	Db	6896	LVAEBARTAAPDLAAGLAARSLDPGEHRLPADPGLVSRRLAEVGVSTGFDWVERLSA	6955
QY	156	NVSK-----	165	QY	204	-----	207
Db	5876	NTGDERNMPVISLDRPADSLSPVAVVGICRFPQGVNSPGFWDLLTAGRNTVGMPPD	5935	Db	6956	GLGVLHAQVLSPDASSWAPLLDAVMSIAPAFVGLPQLRMVVHVDEITVDGTPPEAATVE	7015
QY	166	-----	165	QY	208	-----	207
Db	5936	RWEYRDFGPRDAAALRTAIRSGSFLDDDIAGDAEFFGISPBEALMDPQORLMLEVAV	5995	Db	7016	VALDPRVADTVHALVTDGGRPVASLRGLRYPVVBQPAAPDTDEPGDADADVVSFAGLS	7075
QY	166	-----	165	QY	208	-----	207
Db	5996	QALEHAGIPPHTLAGTDGTGPAGVCTYDYGAGLEDLPNIDAWTGIGAACVAVSNRVS	6055	Db	7076	PEELASRVLDEVREQIAQEMRLAPTALHVRPLVEQGLDSVMTVVVRRLEKRLGRDPA	7135
QY	166	-----	165	QY	208	-----	207
Db	6056	LDLRGFSLSIDTACSASLVALHTAAQSLRGECTVALAGVGNLLVSPQTIAGTAGALA	6115	Db	7136	NIFWKLPTISDIVHLTERLTHEPTADGHASXTHXLRNSCQGTAGXKGRASRPGNLGKT	7195
QY	166	-----	175	QY	208	-----	207
Db	6116	PDGRSKPFAASAGRYAASADGYRGCGVIVIKLLTDAVRDGRVLAVLRGSAFNQD	6175	Db	7196	XKTROYDRCAETPDPAIWSSLETTXSGDNDRKVWDKALILVGGVSRRLRPITHTSAKQL	7255
QY	176	-----	175	QY	208	-----	207
Db	6176	GRTNGIMAPCGQAQEHVMERRALTAAGVAADTVDFVEAHGTGTRLGDPMEIGAIAAVYGRD	6235	Db	7256	VPVANRTVLSVLDISIKEAGISEVGIIVGATAAAEQASVGDGAEFGLDVYIQQDAPRGL	7315
QY	176	-----	175	QY	208	-----	207
Db	7316	ADAVLVSRLFDDDDFVMVYLDGNYVDGIVDPVDFRDRDKPAAQVWLARVARVDRRFGVAE	7375	Db	7316	ADAVLVSRLFDDDDFVMVYLDGNYVDGIVDPVDFRDRDKPAAQVWLARVARVDRRFGVAE	7375

QY	208	-----	207	QY	213	-----RYA-----	215
Db	7376	LDENGRVAVVEKPODPRSDLAIVGYAFSSAVAEVAGIRPSWNELEITDALQWLADH	7435	Db	8456	ARPLRRHALRAPLPRAAALGAARRHRLRLPLPLHLLGQLGRNDRPRRYAVHPVRED	8515
QY	208	-----	207	QY	216	-----	215
Db	7436	GHEVGSTVIDRYWKDVASVTDVLEMMRHLLSLETRIDGEVDEASBELVGRVVVEAGAKIS	7495	Db	8516	HPEPOARGLSDVRWLDPRPSLMRALEQAGFTDVEVLHERATAVECDIVAACKKAGVPSRX	8575
QY	208	-----	207	QY	216	-----QRNCTHSFYLV-----	226
Db	7496	GSRIVGPAVIGPTVVQDSYLGFTFSIEGNCRAVHSEIQYSIVLRCASIEGVARVEDSVI	7555	Db	8576	PCPDEAKGCRTHDHGDDTGFGRGDRAVPTGRLREERGSGSRPGEDRPSQGRRTAHHPLHR	8635
QY	208	-----	207	QY	227	NAMSRNLFRVPK-----	238
Db	7556	GRETEVNAAPRTSTAHLVLGDHSRVQVGXGTGLRRRTTRRXRLTGTCETAARVNGDG	7615	Db	8636	GPALLRLELPRGRXASRPLPRAQGEAGLPGTGPAPRGTPGAARRRVRGRPCRAHRL	8695
QY	208	-----	207	QY	239	-----	238
Db	7616	TSWSSVERASSAPSTCANWSGTAIRRGPCWTSXPTPGTWTWSRSPAGTSPCTATSAMP	7675	Db	8696	RARPOGAVPGRGGAWAPXPRORRPAHGLQSEHLPORRRTAQMPPGGSGIAPAPRRRG	8755
QY	208	-----	207	QY	239	-----	238
Db	7676	GCSRWSPATIWNSTSRPNRTSGRSPTRRRSYVPCWCRASHSCRCAMRAPRPGSCRSP	7735	Db	8756	EGARDTRGSGPGQPRRGARRPAPRPGPERQQLAADHHRVRRPORTAGALSAPP	8815
QY	208	-----	207	QY	239	-----YINGTKLKNT-----	248
Db	7736	TRCTAASRAPGTRTRRSPTTRPRPRAATWSPSPTPAPMGCRCAXEGAATTTGPISS	7795	Db	8816	SARQTSNGRSGLSDEYVHRLELGGTLHVHGPGIGAARRGTRDPGVPVAVAGTRRRD	8875
QY	208	-----	207	QY	249	-----	248
Db	7796	RRRSFRCSPPGCWTATASRCTGMAATSGTGCTPTTCGASGSPNAGCRGRSTTSRAPS	7855	Db	8876	RPDAGPGARRAGHDGKRPAGVLCAGHVHTAAVRSAAAAAAPHRAADAVAOQLRRRRPAR	8935
QY	208	-----	207	QY	249	-----	248
Db	7856	XPIWNXNNCWTWPAPAGTSSVYPTGRTTGATRCATPASARSATHPVCRSSKASPTFC	7915	Db	8936	LLEEVDRGAQLXQGRBELRRPLAARPCGARHHGRGRTGRRAPCBAERLCVARVHRHG	8995
QY	208	-----	207	QY	249	-----MRKLKRQAP-----	258
Db	7916	AGTPRTGTGGSXTSARCPTAMLWPCRHQHPXLGESPPAGFPFADRILCFALFSRFRPD	7975	Db	8996	DRAGARPLGGPAPVVLREVRVRVGRPDVPRGGPLTGRGAAAGRAAAAHALRALQRRP	9055
QY	208	-----	207	QY	259	-----	258
Db	7976	GALSGDRSRLHLACGGSAQPRAPGIGPRRPQRRHRACPRGRAPVHRVGDRCAGRQ	8035	Db	9056	GRGPVAGADPQTRVHRLGQLGLHLRRRTTGAARHRRGGAAGRGGAAGTGGGA	9115
QY	208	-----	207	QY	259	-----	258
Db	8036	ALRRAAIRPLRLGRVVGGSQSLGEEPQLRHQDQREPHORGGAANRRLVLLFRVLRG	8095	Db	9116	GSAAGRRPGAAQLPAGAHPPVLRPAGAPRQQLLHERHRGGDSAAVPRPQLRHADLRPPY	9175
QY	208	-----	207	QY	259	-----	258
Db	8096	VPARRDPDAGVVDPTGTRQLRQETHGRARAGNDADPGALHRIPIYAQRVERMAEHARP	8155	Db	9176	RSRRRHPDPVRPXGHGREGGAAPGAVRPLPPGGGETAGRRRAGTFARRGRRADPPR	9235
QY	208	-----	207	QY	259	-----	258
Db	8156	LSQCGRHFLQDPARXADLGLRRRSAGVQLRQGRHRARPDRDEGLGQLORLVP	8215	Db	9236	GGRAVGERRRGRQRHGRSEERVITIGLKQORAVTGGAGFIGSHLCERLIERGLSVT	9295
QY	208	-----	207	QY	259	-----VKEQFE-----	264
Db	8216	HQHRAGAGPACGGRCPQPHRAPARPGDGLRHLHRHGGPRGLRLLGGHPAGGARPD	8275	Db	9296	CVDNLSTGRIANLDALADEERFTLLRADVTEPFSVEGPHVVHLASPASPLDYLALPLE	9355
QY	208	-----TALLI-----	212	QY	265	-----KKA-----	267
Db	8276	RRVGRLLRPRRTALLLRDRDRRRARPVGVCGRPAERRPVIPSRLAWKRTXQWRXG	8335	Db	9356	TLRVGSAGTENALRLAVAAGARFVVASTSEIYGDPAEHPQSESYGNVNPIGPRSVYDEA	9415
QY	213	-----	212	QY	268	KK-----	269
Db	8336	NSERSXDSASGTXTRAVERLRCVQGRPVHRHLHARLPABPAAPGAGRVLRGVFGQARP	8395	Db	9416	KRFTREALTAAYARTLGADTGIALRFLNSYGPMMRDDGRVVFTFIDQALAGLPLTINGSA	9475
QY	213	-----	212	QY	270	-----	269
Db	8396	GARRAGRGHPGDVGTARRRDPFGGFRREPAPRRVHRHGARRHORGTTPARRRGVDGVR	8455	Db	9476	QTRSLCYVEDTVRGLMALMDSSFPGPVNIGATGEMTVRQIAEIAAALAGVELRTEFRAPA	9535
				QY	270	-----	269

Db	9536	EDEPGRRCDDIETARTKLGWKPEVPLAEBGLKRTLWMWESTYPRDMAHDGVNGSSDIRSGA	9595	Db	10616	PRDAAGRRPAAQHPTARRPGGGQAAQGRLLRVPARRGASARPARARVGAQAAGAA	10675
QY	270	-----	269	QY	306	-----	305
Db	9596	AXQARHPARGCCPGASRNSWMSGGTTPRACRCSRAISPVTSICATVPSGSTSWPRR	9655	Db	10676	RVPGAAGRVGAPGRGALRVPPCGAVPALGRVARDLRPAVPARGVLAVNVLLSRLPLGA	10735
QY	270	-----	9715	QY	306	-----	305
Db	9656	CPXPRTPICRSGAAARASWPASWRPGSAACCACTACSACRWRPAPPVVGPTSPPTTSGT	9715	Db	10736	LAPAPAPGGEPTAQEVIARFRERTDPRPDWAYAHLLDLRNALAEQLAGASGRWLDG	10795
QY	274	TT-----	275	QY	306	-----	325
Db	9716	TTWDPRTPTPDSRSTSSSTXTPSRSTSTCARSGCCGPAAGSSSTTWTARSGRSS	9775	Db	10796	AGTSPYRDLPCARLETAEMRGEDLTADHELDAGLCGLPDGSP-DGVLSTQVLEHVT	10854
QY	276	-----	275	QY	326	-----	331
Db	9776	TSTSRATPGRTRSPSRDCPSGAPTWSASSRRRPGXHYPGRTTTRAASASSPCAGNRRL	9835	Db	10855	PDHLREALRLRPGGRLVLSHGVEHGGQDLRWTDAGLAAQAERAGFTVDRTVKLT	10914
QY	276	-----	275	QY	332	-----	331
Db	9836	REVPHAGTARPGLLRLPGHLGAERVLGQDRAAGRRGPARVPGRQLLRHRLRARGHG	9895	Db	10915	CGPRGLLLRLRYHGREHGWPGGPVGLLLRTTLALADRLRPLVDYLDLVFAGQGRVEGP	10974
QY	276	-----	275	QY	332	-----	335
Db	9896	RGGTGPRPGRPAHPSLRHRLHDEGRGAAGRCGPQORRRFPARQDPALAGDRLRG	9955	Db	10975	AEFFYLDILLTASKPREPEKPERDAIXPSRPVPPVCVCCRGSTCAPSTVSTAAAVTTTR	11034
QY	276	-----	275	QY	336	-----	335
Db	9956	RVPGPLARPEGPAGRDGGCAGVGGRARPVGRQLQHRNGRILLRRHPAGRPGAVOH	10015	Db	11035	CITTSCTNPRRSRSSSTPTLSAARTPATHRTIPMRSEGXPRRSTAPTCGGPGMRCWA	11094
QY	276	-----	275	QY	336	-----	335
Db	10016	ARPRHREAGAAALRRCDRRHGVGARAGGAGSAPSPGVPARRLARLLPDLPGRTLQ	10075	Db	11095	SACPRATTAITARTGARRPTGRASTSGSPTRWTSPIWPNVPPPTARSRSPSSSTSGSAT	11154
QY	276	-----	275	QY	336	-----	335
Db	10076	AARGGPABGVRRRTAQCRAACILGAGPPLGSHFGHRRPASAPGQRPGRPRRG	10135	Db	11155	PATCGAXCAPTRSAAGVPRPSARNRGRPXSTRGTGRTRSCGRRCARTPTSSGGCTSAASP	11214
QY	276	-----	275	QY	336	-----	335
Db	10136	RAARPORTAGRRARAGRQGTARARREVLGRWPRRCPSSTGTTXPSRSSNGSAPP	10195	Db	11215	SCSPATRCWTCXSRPATKSTSTSTPTSRCCRRSSSTASRTSRASGSARPATASRSS	11274
QY	276	-----	275	QY	336	-----	335
Db	10196	TPAAWTSVPAARSSRTCAPPRGACTSRSSRCRSTPNSCAATSPRSRYGRPRPPPLAG	10255	Db	11275	SASAPRSGTSSSRTYAVPRPRXTCGSRSPRSGXTCGDCGTCCSSPARRASTRWPMSCS	11334
QY	276	-----	275	QY	336	-----	335
Db	10256	TASSTWSPATAACAARVTTAPTRPWKRSTWTPYVWTTSSPRTPAWTWSRMTWRAARWA	10315	Db	11335	GPTTCRSPISRTRSDGRPSWPNPSGATAWTGASPPPICGRXARSXTPLTRPNXSPKE	11394
QY	276	-----	303	QY	336	-----	335
Db	10316	PCAGRSCCAGFPWWSSNTAATRCATTAPATTCCRCWSTTSATRCPCRAGWRRSA	10375	Db	11395	EYERVESCPCGCGAAQDLRIRDGAGRVPRRAGRVPRPARSQRQGDHRRTHPEHPAQ	11454
QY	304	ST-----	305	QY	336	-----	335
Db	10376	STGCLSPPCARSSTPIAGPCPPVPPEEKCIHRDYNCLPLRLPRGPRMPYGDPRAP	10435	Db	11455	AGRRPCGGRPCDRRRGAPEHRRHRVPGGGRGDPDRPREPGDVGPLPRPGQGAAP	11514
QY	306	-----	305	QY	336	-----	335
Db	10436	GARHRLPAAGCAGRARPAEPRRTGVGAGGGRPDRERLAHPVHARHLDAGAPRARRGPG	10495	Db	11515	YORTAGAVRDRGRQAQALLGRYAAQAPQRRILHRAQALPGRAHHRPRPAQORGL	11574
QY	306	-----	305	QY	336	-----	335
Db	10496	PHRVHGVPGDGPAGGPAASAAARHCDPGRAELHPLHGQGVQDHRGRAPAGLAPXRPDV	10555	Db	11575	EDGPPDRAGHRAAHAVPGGGRPARPDLDGDRGPRGRGHPRDPEVPGGRRDRGDP	11634
QY	306	-----	305	QY	336	-----	335
Db	10556	LACRBRHRRSRHRRGLPRPARGQCLGHPAVOLLLGGRAGHRRRPAQRHVRGDRGR	10615	Db	11635	GGHRRGVRGSGRPLHGDRAAGRRGPPYRAGTRPGRRRPRGRDRAAAAGAPRRGRCA	11694
QY	306	-----	305	QY	336	-----	344
Db				Db	11695	APPDAGXRVPRDQPCRSKEGAGRMWAVLGRARWAVTDGWWVRANMINWRRNPGLI FYC	11754



Db	13915	DRTAVIGATGFTGSRSLTARLTADGPRAPAFPPVVGRAAPGLAEADIVYFLAAGL	13974	Db	14995	CMSRMARRAPTWRPGRWDGTGRSGRPRGPTPICATPSWCTRRWPTPAPARASWFSRTC	15054
QY	362	-----	361	QY	376	-----	375
Db	13975	SPVLAERRPDLVEAERLLVEVLEALS RAGHRPVFVLASSGGNAVAYAPEVPPYRETSPT	14034	Db	15055	CPWSSNSTGPTGGTPRSSRQCAGAWARTYRDRERAGDEMAEKQAVAPWMTDLRLVFGGM	15114
QY	362	-----	361	QY	376	-----	375
Db	14035	POSVYGHAKLERELCRHSDAVRAVVARLSNVYGPQRALRGFVLPWHLTAAARGEPV	14094	Db	15115	ATQVVGLAVRLPDAIGERTADGLAADFSEPAAMNRLRLGLAALGVLRSEKPGVFA	15174
QY	362	-----	361	QY	376	-----	379
Db	14095	RVFGPHVVDYVHVDVTRFLALRGVGGRLPAVNVGCVFTSLGELLEIVSGVTG	14154	Db	15175	LTPVGELELRADSPSFHSLARMLTDPAVATAWQHLDHSVRTGPGAPDHFVGRDFFAYLAD	15234
QY	362	-----	361	QY	380	-----	387
Db	14155	GSVAVWERGRSFDRCQNWLDVARABAELEDWRAAIPLTEGVRCQWVRLHDAGRAPSAST	14214	Db	15235	DPDLWLWYNAAMSQGTGGIAGLVAHAHQDFSGVRTVVDVGGDGTLLAAVLRAHPSLRGVL	15294
QY	362	-----	361	QY	388	-----	387
Db	14215	RXTVRRVRSDASADGAMPGSTLVPTWXSASTLXSGPGPADSORRHGHOVETRGG	14274	Db	15295	YDSAAAGVAQAGEVLAAAGVADRAVEAGDFFAAVPEGGLYLLKSVVHGWEDERAARILAH	15354
QY	362	-----	361	QY	388	-----	387
Db	14275	VVEVGVHLAGSLADRCGGHPGRRHQTGVLGRREQADGVLRQGVGLVRLGAVLVRA	14334	Db	15355	CRRALPAHGRIVWVEHLLPDTVPADAVPTTYLNDLMLLVNGNGLERTRGDFEOLCAAAGL	15414
QY	362	-----	361	QY	388	-----	387
Db	14335	GDGVVDVGOSRQDFGEGHRAVDLLVLGRLGRHVRVQLEVGGLAGLLLVHAGHQEQPL	14394	Db	15415	TVEAVTLAGTDLWLEBAVPASADPLGXAALPGISPCSGASLRDRAEWSSHAPDASEFT	15474
QY	362	-----	361	QY	388	-----	387
Db	14395	VGVAVERLVERPGLLLRDLVHRGHEGRGQAAVERQDAGQOQVBEAGGVAGEVWAG	14454	Db	15475	ILLARTVIYLGSSRTVSGDWRYALFTAGRXFCLGERSLRESLARMALTSRVNSNPDMS	15534
QY	362	-----	361	QY	388	-----	387
Db	14455	AVEAVGDVGEVLLHPYARLVDAVRHEVAHGRGFLDQGOHRLQGVFPALQSPVLVAD	14514	Db	15535	SEWCAGRAGMPATVGPCTRTSEWXAATCWTTRAXLSLLSSPAPPSTTSAPCRSSRSPG	15594
QY	368	-----	368	QY	388	-----	387
Db	14515	TDVAVPAQDRHPVPAVRGADDVADPCGRTGOORALPVAPDRDVAEDLVRAAVLQVLGD	14574	Db	15595	SSPPLPRXPVAVCCGWPRSPASRGSVPPPPAAASASIRPPPRCSCSGSPRRARP	15654
QY	369	-----	368	QY	388	-----	387
Db	14575	DAGPARGVDDVVELLADQLTGVGQPLRPSRAVAVEQHALDRDALGDVGALLDRVLQEHV	14634	Db	15655	RPSTRAAPTGWPRRSRTTWCRTTASRTGCSAGSGCSTTSHRSAGRSSRARCRCGWPPX	15714
QY	369	-----	368	QY	388	-----	387
Db	14635	VEVAVDLPADVGTEDLRPVDRPVLLAPAQHAVLDRVTAGVDLLQRADEVDDGAARGG	14694	Db	15715	SWAXWTSSGNATWRNRNATGASQGTTCAXRXKCSPTSRRTSSAPGSGPTGCAIWIISPSR	15774
QY	369	-----	368	QY	388	-----	387
Db	14695	DRLADVAREGLPFQXLHGHALEOGGRRGRSPTADHEHFFGHGAPHXPFHFVAHHC	14754	Db	15775	TSQWTGATTWRTPGCCARRRWCNCSAPTXYAALGRSAGSPRSRNRNWSPXWSTATCACSP	15834
QY	369	-----	368	QY	388	-----	387
Db	14755	GKXKTSGTGNCCELLRVXSSRSSGSISSPCATMXHPLRDXCRAPRPSCKGAXWEPAYXP	14814	Db	15835	TVRTPPAPFNSRNVCGEPGPRGRRRTSTSPPLGGVPGVRRRCGVHRLVRPVLRRAV	15894
QY	369	-----	368	QY	388	-----	387
Db	14815	RRTXELRLRLHPLAGCVPRGARRRGVRPAVATGHGPRRPGHLDAGSGPAGGARRRRVP	14874	Db	15895	PGDAGAADRAAAMAVRRRHGKPRVRSLLRRRGPAHQLRAGPRVEPAHHQHPAGDRHP	15954
QY	369	-----	375	QY	388	-----	387
Db	14875	ARRQHPDAARRLRPARRGFVAARXRSARSRCGFFXPSRGRPRRTTACTSTFASWPRTR	14934	Db	15955	RHAGOGGRRRSALPRAGQDRARQPHPAVADRPGHPQQLHPPHPGQPHPLSGVLHQRR	16014
QY	376	-----	375	QY	388	-----	387
Db	14935	TPGTGVPVSWTXSRRTTTSSATTCGPAAGCCSCSRTPARRTPRXSAAPTWCRR	14994	Db	16015	GHPCGRPAVRAGRVVSAQAQPHQGRPGXGRARQGLPLAAADRATAAAAHRRPQGHG	16074
QY	376	-----	375	QY	388	-----	387
				Db	16075	PDRAGLPPGGYVPRILPAGAGRXRRGGTAVVDGPGHLPERHGGRAALVHRGOEPAPDTGT	16134

QY	388	-----	387	QY	397	-----	396
Db	16135	PDAPAGPARVARHSGADHPGRGAALQHHRGDRPGRXPRGGRMGPAANVPAGARSRRVDO	16194	Db	17215	SVTSIRCATTERRCRSGPGRRASTTEATGRAKEVGGRRIRPLPSLPMWRKSRASGKXGSAS	17274
QY	388	-----	387	QY	397	-----	400
Db	16195	GHRRGAPARPRVPAPRAGRHGDGPDRAVRPGELSAARAAVRLRPDRFPGLQCRP	16254	Db	17275	TFRNRTRSNRSGSGSARTPIITRPARSRGRKCCATATRIINCGSRPGSNRTVEPAD	17334
QY	388	-----	387	QY	401	-----	400
Db	16255	GRGGRALLPLAEQLSPGRRRLPRSRPXLLIGDRPPAHAPAARLLRQRGPEPARLP	16314	Db	17335	LGLQVGLAEGPPRADVERAGVRSAADRQRRASAPAVDATPARVACGNVARGAGARW	17394
QY	388	-----	387	QY	401	-----	400
Db	16315	AELVTPXRTSPCGWAFSAAPSPDAVPCRPXRCRXXPPSPPCARTTPVRSRGSAA	16374	Db	17395	GGVSGRRWPGAGFTTTTTTHAPSRPYGRPMNRSMGNWPGASWRSRWPGCRTPS	17454
QY	388	-----	387	QY	401	-----	400
Db	16375	RPSSATRCWRARMSTRSTSPCPDCTTTGSCAPCARESTCSRSRXPPATRTPKRRSAS	16434	Db	17455	RAGCAGWRRRAGSRGSRARCSCTPARTGCASSAGSGSRPTTRWPPPSGSPRLYGTASPYT	17514
QY	388	-----	387	QY	401	-----	400
Db	16435	RTPAAXXWRTXPSPTCTTPCARCWTRAASVNCAPSPASSGRHCLPATSGTSRSWPA	16494	Db	17515	STWTGSPGPPHWSRSCCTWRRAPRHAGPSRRAGXRRRSARCWPRRRRQAGCACCSR	17574
QY	388	-----	387	QY	401	-----	400
Db	16495	AACWTRACIRSPRACSSGPTSMWGRRCGWEPKXSTWRATCCVPTAARHTCPSASS	16554	Db	17575	GGPHGDGRWRPWRSSAHXPSRPTTTVCHRCSPRPRSTCCAAARAPRWPWMWSSCAPPSTT	17634
QY	388	-----	387	QY	401	-----	406
Db	16555	TPTAARTRCGARAGSSSTVSPSRPTSAFRFGWNAPTACASACRPTTSSPAPSASSG	16614	Db	17635	ACSPSRAAPTPRRPRGSCRTRWPSGSPRRSTTRTAGSHCAATRPAPASPSBSAGRS	17694
QY	388	-----	387	QY	407	-----	406
Db	16615	PSPAAGRRTPAPMSHGPDWVCRAGHGCATAGOEQADRRERVLVAGTGFVGRQL	16674	Db	17695	SATRWGARSRRSPSPHSRTXPPRGRATSWWSVRPACXSTGPRPSRVSACAXHRTAPTG	17754
QY	388	-----	387	QY	407	-----	406
Db	16675	CADLTAAAEVAIAARVPDFPLPGRLLVDVTASPGELADVIDSVRPHVTVAIGSNW	16734	Db	17755	RTRSPTRGGAASRPPSTTRCGASRRPGAAWACSSWTRPTTYRIPGPAAPWPSPSGRST	17814
QY	388	-----	387	QY	407	-----	406
Db	16735	GIAERDMETNCAVPTRRLLDALRRTACRPYVHVHLSVLEYGPTPPGPEPTRTASPPRTTA	16794	Db	17815	ATACCSXPVRPWRPTASRASAACWCGSSSPSWRRPSTTTAWRAPRSARRSHSICAATSR	17874
QY	388	-----	387	QY	407	-----	406
Db	16795	YGAKLAASQAVLEAAABGVEAGVLRIANVAGPGTPANVLLGRVAGRLAEAVTRDWLPA	16854	Db	17875	TSSPNSRRSSTRTSGRSPARRTRPTARPCADATSWRCAGRETRVRSRPRCTGCARSLR	17934
QY	388	-----	396	QY	407	-----	406
Db	16855	VVELSQLRAHRDYYVDRDVSDAVLAATRAIRPGLIWPVIGRGEAVPVRWLVDLLVEVSGVP	16914	Db	17935	RPRTGTXRSWCSPRSGTYWQWYARRSRTPOGRKQARTSRGPKRGEQPMRGARRWKG	17994
QY	397	-----	396	QY	407	-----	406
Db	16915	AEVRELPAATGAAGDDWIOVDFGPARELLGWTAVRSLHESVTGLWADTLRROAIPQHA	16974	Db	17995	QPRTERARRRPPYGTHERRRSGRGRACDRDGPVCVSGSRSAACRPPADSSSTSP	18054
QY	397	-----	396	QY	407	-----	406
Db	16975	PRFSXALSSRLPRRPPSGRTTGIRROGPAGGAGSQTPETGPHHSTRPAPOGVTAAAR	17034	Db	18055	RPVTPYCSRRRQAEWASTCRPPPMXSSASRSSRPSNTRRSFVPTAWARSASACTACS	18114
QY	397	-----	396	QY	407	-----	406
Db	17035	HASAYARTSGRLPSSRRRANRIGESPRTRVVTTPRLISGALPVATROMPSPATRGFS	17094	Db	18115	ARKGWTNACWCWRTSPACSTLTTPAAARSFRRHTPSTRTSPWPATSWRNRHGWGRF	18174
QY	397	-----	396	QY	407	-----	406
Db	17095	AYRGRCALRPGASRRRTAERSSSWRACPRGSRTRSPTSASRCXSSSRVPSAPCRSRDGP	17154	Db	18175	RSPLFGGGGBGPPGGADLEGGAVFGRVGLVADVGVGAEGNLDALAAAVARLEP	18234
QY	397	-----	396	QY	407	-----	406
Db	17155	GPARTAACTRRXGAPASSPAATAPSPATRRRRRXXSWPRWRTTRPCGCRWATTGSLPC	17214	Db	18235	LKCGAHLSPTRRSRLFDSCGVMACDRSIAYLRARMDTSSGPETHWLPRWLSVXARCWKS	18294
QY	397	-----	396	QY	407	-----	406

Db 18295 GVSRRVKGPSACAEVRMGRNWRNRSFSGSAHSRRWRSRMTFGPPSSMYSGSHTMKL 18354  
QY 407 -----WGFO----- 410  
Db 18355 IIGGHLRCRISWRMRPSVLASSSSSSGGTGRSRTARAXSRVWSRPGRTALSYQHREMA 18414  
QY 411 ----- 410  
Db 18415 SRSCTMYCSANGSNWISGMFSTPAANVPFVSRSSSSAWASSSRGRLPCSIDAMVSSSA 18474  
QY 411 -----RTFIPL----- 417  
Db 18475 RTALPRELCVKPALRAACICLTIRMVDRFFPRGSLFLEMLAXLPTRTHKDVSTRTKS 18534  
QY 418 -----WD 419  
Db 18535 EYGCTHXKPIILDIRPVNDANELPHQDOFYARNRISIPAAARNLTSMSLKYWGSLDWD 18594  
QY 420 YLDSLL-----FLDEIR----- 431  
Db 18595 RADDVLLCVSELATNALLHGVPFGHGLLRVRYDGDVLRVEVHDSGTGVPRTDDPDEGG 18654  
QY 432 ----- 431  
Db 18655 RGLLVGALADKWVGERTPGKVWAEPVQGPCAIPVEASIFRLTHRRGTVSXHAPHS 18714  
QY 432 ----- 431  
Db 18715 AASPPOAHLRLRTPQMAVNRHTRRLAALSVTATAALAASLVSALEPATAPEGRIOYAG 18774  
QY 432 ----- 431  
Db 18775 AADAVADSVTVNLKADHARSGSEAGRALVEKYGADIERTYKKALNGVATEASEABAKALA 18834  
QY 432 ----- 431  
Db 18835 ADPAVASVQNRMFSDATQTNPPSWGLDRVDQKNLPLNSSYTPDSAGQGTAYVIDTG 18894  
QY 432 ----- 431  
Db 18895 VRITHSDFGRASGYDAIDNDNTAQDGHGHGTHVAGTVAGNAYGAKKIVGVRLNN 18954  
QY 432 ----- 431  
Db 18955 SGQGTAAQVAGIDWVARNVAVKPAVANMSIGGADTALDTAVENAMASGVTFAVAAGNES 19014  
QY 432 -NPSLRSP----- 438  
Db 19015 TNASTRSPARVTEAITVGATTSSDAKAGYSNYGSLDLFAPGSSITSAMWSSDSATNTIS 19074  
QY 439 ----- 438  
Db 19075 GTSMATPHVAGAAALHLAANPSATPSQVATALTSAATTGVVTPGSGPNRLLYGVRHD 19134  
QY 439 -----TYVN----- 442  
Db 19135 HPPGPRFENTGDTISDNSTVESPTVTVSGVGNAPSALAVEVHI VHTYIGDLQVLQIAPD 19194  
QY 443 ----- 442  
Db 19195 GTAYTLKSYGTGSSDNINTTYSNVASSEAAAGTWKLRVSDNANYDTGRIDAWALQFXPS 19254  
QY 443 ----- 442  
Db 19255 LTPSPRPGTAPGDPANREDDPRPAPARRLRQSGRGLPGVGTGCLPPSPSRGCPQLR 19314  
QY 443 ----- 442  
Db 19315 MLRRWGAYHPGPAVAYGKGTCHGVSARDRROHAAPGPRVPPVGRTPPGAAPTHNAT 19374  
QY 443 ----- 442

Db 19375 QGYAHPPPTGRRRHAHAHPHRRGITTGVPISRLAPRLPFAQSYASSRRPALPPEPG 19434  
QY 443 -----LTFPE-----HRR----- 450  
Db 19435 ARYPCYLAPPDPNSRKGPRSSPKGTRRSTRHRRHTAHFPGPRDRTTTTPTAPTAP 19494  
QY 451 ----- 450  
Db 19495 TAAPHPSTAPSPAPWSSASSAFCRPXGMCWDXSRWRRSRGASEARAWRXPVRCSPASDS 19554  
QY 451 ----- 450  
Db 19555 SCGRCCPRASPPTSTASTRTPAARAPPTPSPRATASRTPTWRPTTTWTRCPAPPR 19614  
QY 451 ----- 450  
Db 19615 TTARCSRSSPSPAAPSRATTNSPGPRTSVTRSPPLRHGHLGRTGRRGLLLHPDPPELA 19674  
QY 451 ----- 450  
Db 19675 LRRPRDLFVRRHAGRRQADRLTAQRDDTTRRRPGRLPLRDQRRHRAVRGAGGIRGRP 19734  
QY 451 ----- 450  
Db 19735 GREQGLGGRPRRRARRADRGLRHSWPKGAEPVADLVEEMEDARKDWAKAAAATNADTY 19794  
QY 451 -----AVNLST----- 457  
Db 19795 YTYDSGYEYVDGPATVTARKALGLDITVPSYEEDPGTSGSGSGGGGLNVKAEVXAR 19854  
QY 458 -----NSLW 461  
Db 19855 LXRGESACVKPSARTSHHIEFSGLCGLHGTGTCQAVAVCTSW 19898  
RESULT 12  
ABP76681  
ID ABP76681 standard; protein; 19938 AA.  
XX  
AC ABP76681;  
XX  
DT 26-FEB-2003 (first entry)  
XX  
DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 5.  
KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
XX  
OS Streptomyces viridochromogenes.  
XX  
PN WO200268436-A1.  
XX  
PD 06-SEP-2002.  
XX  
PF 24-AUG-2001; 2001WO-BP009815.  
XX  
PR 25-FEB-2001; 2001DE-01009166.  
XX  
PA (COMB-) COMBINATURE BIOPHARM AG.  
XX  
PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
XX  
DR WPI; 2003-018650/01.  
DR N-PSDB; ABZ37516.  
XX  
XX New avilamycin derivatives, useful for treatment of infections, and  
PT nucleic acid encoding avilamycin synthesis enzymes.  
XX  
PS Example 1; Page 68-301; 319pp; German.  
XX  
CC The invention relates to avilamycin derivatives (I) with antibacterial,  
CC virucide, protozoacide and fungicide activity. (I) are useful for  
CC treatment of infections (bacterial, viral, protozoal or fungal), in human

CC or veterinary medicine, particularly where caused by *Staphylococcus aureus*. (I) are more hydrophilic than known avilamycins. The present CC sequence is that of an avilamycin synthesis enzyme from the *Streptomyces viridochromogenes* Avilamycin A biosynthetic gene cluster (AB237515-AB237516)

XX Sequence 19938 AA;  
SQ Query Match 21.9%; Score 536; DB 6; Length 19938;  
Best Local Similarity 1.2%; Pred. No. 0.011;  
Matches 218; Conservative 96; Mismatches 135; Indels 17947; Gaps 66;  
QY 1 MGRKE----- 5  
Db 4 LGREAGVRWQXTVDLLAEITRDAQVHRLLTESHPLPSACTDSNLSLATLGCVPQAKARE 63  
QY 6 -----MMVR----- 9  
Db 64 SLDVNTCAGGLYAGTTSPLXPRSHLSHTEAAAAALAPAVAGAGVLLVRRHGGVQAQS 123  
QY 10 -----DVPKMFVLISIFLLVSFINCKVM----- 33  
Db 124 LARGDGRGADV-----LVSAVVVRVVRVGRRLRPVLPGLVLLDQIGD 171  
QY 34 -----SKALY----- 38  
Db 172 RLGALRPVSAQPSICSPSTAMTSPAALFAARSSAYSSSYSASTALVAERKATW 231  
QY 39 ----- 38  
Db 232 SASVVGSLRSEPVSLPSSCVAPNRQVIRSPKQLWRVGMKXXTSASSGTAQVSMAXRR 291  
QY 39 -----NRP 41  
Db 292 RARNTCPRSGXVRRRGGXRPRTPRRARSRAPRPRRPPPCPSRSSRP 351  
QY 42 WRG----- 44  
Db 352 WRGRCGPRRRPXSRRRRRXSPWTATPSTGRVRRRAPNRLSPCCLARPASXAPA 411  
QY 45 ----- 44  
Db 412 RSVAPALPAESRGCGPGRWRRRRVRRGGRGCGSRGGTGRTVPGPRVRAGMATV 471  
QY 45 ----- 44  
Db 472 AGTGPASALGAAPTLLAAVGVGAETVTRVPCSWFCRQRAAGTGVRLGEXKAGGACR 531  
QY 45 ----- 44  
Db 532 NGDPEGDPVAVGVVRMPAAACGGWGVVRVALGCVVGRAGGCGPSDRRHFWARRCV 591  
QY 45 ----- 44  
Db 592 LTPVTAGHPTRPLPAVRDCGPGVRAFSQAHXIGTTPARRANGEAPRTHPAQASTP 651  
QY 45 ----- 44  
Db 652 LPEAACRGTWVLA VGRVTRCRAARTTRTGGGLELQRPGVYAPGVVVRVVAHPQPR 711  
QY 45 -----DQLKLEI----- 62  
Db 712 AVGLRGGVHRVGVVDVVRTAAGAVLERVRRAVGGDQDLEAVDRVDDVDLYEGEGR 771  
QY 63 ----- 62  
Db 772 VAGDADRHRGLDRGVADVVTGVLEARARVVVPADVQBPVGRARAGVDDAGGGG 831  
QY 63 ----- 62  
Db 832 RORRGLGRGCGVGEVERGGARHVRGRHRTGTGVRVAXVPGRGDGTGSEVBDG 891  
QY 63 ----- 62

Db 892 AVTVAGLVGRARRDRDGLDACGKPRGGVGRVLVAGRHEGDAAGHGVA YGRVEGRVG 951  
QY 63 ----- 62  
Db 952 AAAEGHVGDGRLDGVDPDADGDDLGGGALAGVVQHADAYDLGLLGDVAGVADGRAGHV 1011  
QY 63 ----- 62  
Db 1012 RAVAVAVLGVVVVVDGVAVGGPAAEVAVGDADAGVDDVRRHSLPGRVVRGVRAVQGEVL 1071  
QY 63 ----- 62  
Db 1072 LVHAVQAPGGRVGLGRVDAEHAHLDHGGDRRVRGGERLGLRLGLLDRVAVOGLLVRLDVG 1131  
QY 63 -----LROLETTI----- 70  
Db 1132 AVLLDEGASGLGTRTGVRLEHVDVAVGHGVCRARVLDLTALGGRGSGESDEAGEGRGG 1191  
QY 71 ----- 70  
Db 1192 RHAEQGSPPGMTHHCHLRGPPQSVVRLWGRGGVGRGMSTHCASPVROPKDXGLHRNCT 1251  
QY 71 ----- 70  
Db 1252 RALHGEGLPDDFAGRS LAHAPLVGQADQEQTAAPFVGVVGGSGYARS AVVDLDAEDVAV 1311  
QY 71 -----STKYNVSKQPV----- 81  
Db 1312 IADAKETMAGRDVAEEGVGRQLAHAEYVGVFPVGEAPVLEXPGGQVAGSGDGSAVAC 1371  
QY 82 -----KNLT----- 85  
Db 1372 VELVALVGFVVCVVGTVNAQRDWLLSVRAPVLGCTGHRHVLMGACBELGQHELQEA 1431  
QY 86 -----MNTFFQYXILAGP----- 99  
Db 1432 EEPVVDHADGQADAGRQGGGLHAEPFGQ---SGPRRGDHRVDRTGEAAPAATRPCA 1488  
QY 100 ----- 99  
Db 1489 GRTPRHEGHVCGRREHAGDRPVPAVRGAVHARAGGHLVPLVRQRRPARPAPDRLRPR 1548  
QY 100 ----- 99  
Db 1549 GPARTAVRGRAGSQDGGSHRTPGDPAQVPADDELHRVAGAVHRTGAESEPPRPAP 1608  
QY 100 ----- 99  
Db 1609 PPTRVRRTPGARPPVSTHAHLRTRRTGRPLHPPGDARSAPGLHREPTWQPVGFWPGRGV 1668  
QY 100 -----IQNYSITY----- 107  
Db 1669 HPGAQVCDAAVTGHDHDSAGVEPARPSGGRXMXSTALEWFKSSYSGSEGGQCVEVALCPHTI 1728  
QY 108 ----- 107  
Db 1729 HIRDSKNTPEDGPTLQVSPATAWRAFTSATTEXARRPELTPTPTVPPPPRCGGPGRPCR 1788  
QY 108 ----- 107  
Db 1789 RRPVSTRPRCGGRKRRTGTCSPASAPGVRRPPLPCRAGGARGSPGPGCGHGRPPGVR 1848  
QY 108 ----- 107  
Db 1849 RSAXAAARRGSPRRRPACXGFLRLPR SARAVRRDRARRRRRRAAVGGRARACPRSRARTH 1908  
QY 108 -----LWDF-----YSTQLRRKPAK----- 122  
Db 1909 ARPAPAPCAAPSAPSAPVCPVRRRSPGPACPAWLFPFPPGPAHPLWLAPLRTPARPAG 1968  
QY 123 ----- 122



Db 1969 LLPLARPRXPVPLVVRPGTTRTPRSARSRPPQBSRAAGAPSRTPADARTPPCAS 2028  
QY 123 ----- 122  
Db 2029 PXSCRAARPRGRPPPCWAPPTRPCAGAPGVXGRGAGCAGDRPVVRPCGTPWSPRRR 2088  
QY 123 ----- -VY- ----- 125  
Db 2089 ARWPPPARAGGAPCGTFRGSPWAYRSGTARGRSAPTRRRPWSGGSGDSRSGPRP 2148  
QY 126 ----- SOYNTAK ----- 133  
Db 2149 RRACPPRRPAPGSPAARRWXPRRPAPSSRTRPAGRGAAVSAAHCAHSRSPSSXS 2208  
QY 134 ----- 133  
Db 2209 ARWPDPPXGSAAPRLGARVRWPGSSCAPSRRRSGSGAGRGSGCRRPGSRAVRPG 2268  
QY 134 ----- 133  
Db 2269 ACRAFGSPARPPRPAFTPRPPRGRGRPARRAGRSTRRSAAARSGRPPRSPGGAAG 2328  
QY 134 ----- 133  
Db 2329 RPRGAAVAHRRGRPARSVRSGPPRPPAPVVRPAPRTACAAGLPPAPPRPAGRASS 2388  
QY 134 ----- 133  
Db 2389 APRPPRRPGVPRGPPPARAARPPVWSGARACPRCVRCCPVQPRRSAGRRPPSGRRS 2448  
QY 134 ----- 133  
Db 2449 AARPRGAGAAGTSRRRAPGRSGTRFSPPPPGAACPRXGTPAGPGXPPARAGFVAHGSV 2508  
QY 134 ----- 133  
Db 2509 PGPPVPRRRVRDRPPPPAPPREAARPPPPPAHSAPSRPPRSHRVPASCHPGGGRIDCW 2568  
QY 134 ----- TTF- ----- 137  
Db 2569 CAGCPATLTIGRANTCPLVVGSRGALQADRLKAEIGRLNGFVRSPAPAVPDRSRA 2628  
QY 138 ----- 137  
Db 2629 ASXPTTGPGRAGDRGPGGAXRSVPKSRPPVPAWRRCRALLPRRSALPPWKRGXGADSA 2688  
QY 138 ----- 137  
Db 2689 PPNLFCAGLSRRSPSPWSASPTSCRCAPDRGRAROXPRRRPTAAAAGSRPPPARPG 2748  
QY 138 ----- 137  
Db 2749 SAPPSPGRAGSRRRRRSRSGSPSGTGRCPCWPRPPVPSRTARCGRNPARRSAPXCRS 2808  
QY 138 ----- 137  
Db 2809 RRPRAPRTGPGRTSFCTPRSPWSQSASASVREAPGRARHLPRGDGQRPADQPR 2868  
QY 138 ----- 137  
Db 2869 RGDDPSGTRALABFVRPGVAPGGROASAGAVRGVPPGGGDLRGTTGCGVVRSLGRL 2928  
QY 138 ----- RDP- ----- 140  
Db 2929 ALRAARWALSSNRRPPGGRAARQTGXROGLRPRLVLGDCLPAQVRGPQPGHRLVQR 2988  
QY 141 ----- 140  
Db 2989 ADGPAEFAGPGVDLDVPPVGRARGGKLPDLGGAAHLDQIDEPHRRHGLAASDG 3048  
QY 141 ----- 140  
Db 3049 DDQPRNTGAGGGQHRVRDVAHVAVGAQLGQHLDRRHVPGDLGEPFGHPAEQADRR 3108

QY 141 ----- 140  
Db 3109 GARPGHVGDPOHARLHDALRGLOHGLARREBGLAVGRRRPGRRRPGRFTGRGRAVLOH 3168  
QY 141 ----- 140  
Db 3169 XAEVDDVGTARGPPORVQQAAGRHRVGLHVFPFDPVVAADGVDRHRVADGVHDVQFPG 3228  
QY 141 ----- 140  
Db 3229 RRGDQVDEEPARQCEVRYPPGDRHLGSGRGQVRAQLTAHEPRPTSHENALALIGHLLL 3288  
QY 141 ----- 140  
Db 3289 LTSRCRAAVPGPAPDLPIGPVEXHRRCAPPTRRRXRPGHSTGRCRTGRRPAGRARAPRR 3348  
QY 141 ----- PCGR- ----- 144  
Db 3349 RVPAEPWGGSRAGAXRHGRXSGPRSPACTRSGRRARSRRTRVPCGRRHAHTAARPPRR 3408  
QY 145 ----- 144  
Db 3409 LSPRPSPAASPRPHRGARGTCSPPREADTRPRPAGCRRPAPAGTGRCREAVAGTRTRMXG 3468  
QY 145 ----- 144  
Db 3469 SAVHRCRPRPASRRARRAASAGXRSGPSASGRSGPRGGPPSRCPGSGRSAASPRARAPP 3528  
QY 145 ----- 144  
Db 3529 RAGRARSRGAVRWAGRRRPRRHGSPAAPRSRRRPRRTARRPHRCRCPRSGRRRSPSA 3588  
QY 145 ----- 144  
Db 3589 PAPSARPAGYGVGRCSRRECPAARDARSSWGHLQCRQASRLRASTLTXXPCRRSRVSW 3648  
QY 145 ----- 144  
Db 3649 AVTQXSGTATGKSSASTREXLFCEWXKRPSPSARTALXRRRSGEAVRTXSRNRGGPC 3708  
QY 145 ----- -VPSMT- ----- 149  
Db 3709 GXPSGTHCTVGPISMPSSRPGNRAWTSCATPSTMTLTNATTPRCGNSSGWSHATS 3768  
QY 150 ----- 149  
Db 3769 SPTWAVTPMMLKCRPSSWVICSGVPRHPCRSRRGVRCVSWCRLLASVNQSTAPMSFRE 3828  
QY 150 ----- 149  
Db 3829 VFWPVDSTSAAPTSTGTGREAGNVAGROASTVRVSMLTRPSMRSSRRSSVSGSQWXP 3888  
QY 150 ----- 149  
Db 3889 LAAGTSATWTTMLRLRLRNHAPCSDCTSTRVPAPALVKNRXRVLRPWWMRVXLLR 3948  
QY 150 ----- 149  
Db 3949 VAVTVGDSWRLWLPGLSILACTRQCCTPSTSLTSMRMPISGWLMMGLHSGTGPLXVRT 4008  
QY 150 ----- 149  
Db 4009 RRPSTAKKAPDSCARFPVSASNRHRRSSVSGARVTRHCAAEHRAQMSMSAPATSDRYA 4068  
QY 150 ----- 149  
Db 4069 SEGRAPARARPTWPAGRLTTDRVRRPRRRPHCRXAGACRPSWRPVPRTRPASS 4128  
QY 150 ----- 149  
Db 4129 RASAEGRSRRRAVPPPPGAAAGSPPPRGRASPARRRRPPDRAPGXSRCRGRGPPA 4188

QY 150 ----- 149  
Db 4189 RPRASSCRARSPRLSPSRVPPPGCPAGACPSGASRRPPRATPCAXRTPGSGVTRGX 4248  
QY 150 ----- 149  
Db 4249 TGSXPRTSRTTRTSSAPCSARTPWPFRRCAGAGRAWPRBAGTPAPFTSPRWRPDSAG 4308  
QY 150 ----- 149  
Db 4309 RRCWNGNCXAPXCGWPPASAAHPCSPRQRWAASRRPGRRTGRRGCRSRWGRTARQQ 4368  
QY 150 ----- 149  
Db 4369 RSGSGPASRRSPLGCPGAGADRRGHTAGPARAPFRRLHIRIKXVDARGKCHPGQFPXX 4428  
QY 150 -----CLS----- 152  
|:|  
Db 4429 ALAXTESSPSREXCVSPVTGNGTGRAQVNHCSRKQYSEFTSPCWMGRPPGPIPOXGSRA 4488  
QY 153 ----- 152  
Db 4489 RXNSRGRGAERVSRSRGHGLDQOVGSGQRRDRFRDQARRRABLFEVATRALQPVADQQ 4548  
QY 153 -EMLNYSKRN----- 161  
|:| :|:  
Db 4549 VEVVEGRRHGVRHVGQQVLDHDDPAVRRQCPAAVRQNPGRGPLVLPAMDDALEQVEVT 4608  
QY 162 ----- 161  
Db 4609 ALGYGEEVAGLHRAVDARRQHLPLCLRDPGRRVVQHAAQGVRAHQRRQORVPAPDV 4668  
QY 162 ----- 161  
Db 4669 HHGADAGEVLVRRDQAGDPTGALAHRRVVEPRQVGVVRQVREVPVPAEDVVEGRAAGADAV 4728  
QY 162 ----- 161  
Db 4729 VQVLPGRDRGIGEHPQXVEGRRVAVGATQFPDRCOGEHAGLGAQHSERGQAQPVHG 4788  
QY 162 ----- 161  
Db 4789 SGFALEVGGQVGGPFALPDRVGPQPHRQAHDLGHSABHQPQVRPGRHGLLRHLV 4848  
QY 162 ----- 161  
Db 4849 TGELVAVCPGPGALPARPGCTARRAGRVRAAPRAAGSAGXPXSGGCRWPPACAPGRG 4908  
QY 162 ----- 161  
Db 4909 GTDRRPRPSFWPTXAAAGVPPPPRPPGRCPPRHPTAPAAARPGGSRRGSAWARPPRCRPT 4968  
QY 162 ----- 161  
Db 4969 GAGAAAPGRGTRTCGRSCRSPAGSPAHRTPSGRRPRPRPGSWRRRCASRSPRAASS 5028  
QY 162 ----- 161  
Db 5029 RETAPGCTPASAGCHRPSPASWSXAAACSVGLAARNTSSRTPSRRTTSRVQVAGSSGS 5088  
QY 162 ----- 161  
Db 5089 MSSCRSRAHASSAGNAGNPKWKPSETKLSSAPRSVRGFPFSFTGRSRSSALIPKR 5148  
QY 162 ----- 161  
Db 5149 MLHSRAXRRTRDSTRSLHSEKIFTITRAGCSFLSAMRRDVKPELSVGSMSKKVLVIG 5208  
QY 162 ----- 161  
Db 5209 GGPAGSTASLLSKAGMSVKLLERETFPYRHIGESIASSCRTIIDLVGALEEVDSRGYTV 5268  
QY 162 ----- 161

Db 5269 KNGVLLRWCKEDWAIDWPKIFGPDVRSWOVDDPDHVLKNAVKQGADVTEGTVTKRVL 5328  
QY 162 -----DTGE----- 165  
|:|  
Db 5329 FGDRAAGAELWDPDTGELVSEFDYVIDASGRAGVISKHLKNRPHRPFIRNVAIWGYWQ 5388  
QY 166 ----- 165  
Db 5389 GGSLLPSSPTGGINVISAPDGHVWIPLRGDRYSVGVFVCHQDRFLERREHADLEAMLAS 5448  
QY 166 ----- 165  
Db 5449 LVEESPTVRDLMANGVYQPGVRVEQDFSYVADSFHGGPGYLAGDAACFLDPLLLSTGVHLA 5508  
QY 166 ----- 165  
Db 5509 LYSGLMAATSVLATVNKDVTEBEEAGAFYESLYRNAYQRLFTLVSGVYQQQAGKAAYFELA 5568  
QY 166 ----- 165  
Db 5569 DAMTPERATAEYEQVDGAVAFAPAEVLVAGLADIHDAVTGTHEDRAQQAHAQADALPEDNSVR 5628  
QY 166 ----- 165  
Db 5629 QLLAAAEENARLMAAAGMPAAPVSEAPGKMDAHDLYDPATGLYLVTSPTLIGIGRSRXPQR 5688  
QY 166 -----OGC----- 168  
|:|  
Db 5689 XCRASPLGGEYGARHRSVRGCVXRYXPSNASDRLARRGGPGAGVVEHTLPALPTLQO 5748  
QY 169 ----- 168  
Db 5749 GDGGPVEFGACDIIQPVALPVEGSPPLPAHRDGAAGHTGHDLOQAQSRGHARADVHI 5808  
QY 169 ----- 168  
Db 5809 GQQAAPAAQAEQEPGDVDVDIVPDHMRITAEHPHGLAPGRRQVVRQHTPEAQALAS 5868  
QY 169 ----- 168  
Db 5869 GTVDIAQPRDHTDGI RVPQAQLAFABELGV PVDGVGPGRAGLPVRRFDLGRVDRPARAQO 5928  
QY 169 ----- 168  
Db 5929 YEHRVSRSGORLQHLDOQSAFGLDQVGTFCGEDRTEAGQEVDDVGLRQSRGPAADHR 5988  
QY 169 ----- 168  
Db 5989 RERPVEGGRRARAVSCEPGRQPGSDETRRPDHRRTVPGQVRRRGHSAPRRARGDAARSAR 6048  
QY 169 ----- 168  
Db 6049 EATRGTSPPGGTVTARRARXARVRSGWTPGGLRXPSSGSTNTAGQPRERAGATSWSTEL 6108  
QY 169 ----- 168  
Db 6109 PTTSSSPGGRPSRARMPSSRSGSLATPSKQSRITTVSRRTAVSPHSASTSAHSGXPATTV 6168  
QY 169 ----- 168  
Db 6169 VGISPARAVTAIEPGISWASSRQARRTRTQLSSAAABERMPPRSAGVPPSRXKRSVSSM 6228  
QY 169 ----- 168  
Db 6229 KSRRVTILGSAVTARSKVPXASQASSKSSXXMTPTPSNAIALCPVIATRSRILHTSRREY 6288  
QY 169 ----- 168  
Db 6289 TRINGSAARRPWIRHYRQESSQFTILRLKPMPSGEEIRVHLKCLTKGAAVWGPWKXLST 6348  
QY 169 -----GNFTT----- 173  
|:|

Db	6349	RSTAQSPRTD	SHRCRGSWTR	SWSCGCAVSTIACSPSTSLRATTSTSPAARIPTGRPG	6408
QY	174	-----	FNPMFFVY	-----	181
Db	6409	YRSCARRNSS	PSWSRVRISSGAGRS	PRSCWTSRRTSSTSMGTPPXSRRTTARQRCGTRT	6468
QY	182	-----	PRWN	-----	185
Db	6469	RATWTRAGAGS	AYGPRWTRPWTAACTMCPARTSVPCCTGTSTTTTVSGWXPTRW	6528	
QY	186	-----	TKLYVGPTK	-----	194
Db	6529	TRPPGCPSPSR	VRVPCWCTPCALRITRAPISFTRRAGPMCSWGRPKRSPTPSRACGWTR	6588	
QY	195	-----	-----	-----	194
Db	6589	TDDVPRPGRSTR	SGRPLTSRSLRPAACGTAPHTVRRSHGPTTGPAGHPRRMPAALPG	6648	
QY	195	-----	-----	-----	194
Db	6649	SGVLRARSRCR	IPRLRRPARSRPGSRCPCPRRCSRPRRPPSHPARTRCPTARXP	6708	
QY	195	-----	-----	-----	194
Db	6709	PGTWPAPVRVR	RAGRAAAARPPHRSARRSRTGEXSRRSTPAPRRARPAPGACRRRARTR	6768	
QY	195	-----	-----	-----	194
Db	6769	PPAPPCAARPA	PXRAGSPWPVPGPRRTGAPGSGSRARRRRPRSGRAAPGNCRYSP	6828	
QY	195	-----	-----	-----	194
Db	6829	RGXPAWXS	GXXRCPAAAPGRXHRRCGRPRRRPSAATARSPPAPRPPARRSAPGV	6888	
QY	195	-----	-----	-----	194
Db	6889	RPCPCHPAGR	PRGRVPAARPSGWRPRTGCRRTIPVGVRSSRHSRSPGACXPAPPSTAS	6948	
QY	195	-----	-----	-----	194
Db	6949	SGRPLRAAPAA	SQAVVPAAGPHGRHPVGEPAQRLDGLPRSGRGRDAADDAAEVVRH	7008	
QY	195	-----	-----	-----	194
Db	7009	EPAPGPGGLR	VPVDVGEVVAVAPHVDPAVRAGHRPAAAPVLDQAQVVOGHPARP	7068	
QY	195	-----	-----	-----	194
Db	7069	AGPAAQVGLL	VGVQHLVEQDLAQHRGPGCEERTLIHIGVLQRPWGPPLRQPRRGGKGP	7128	
QY	195	-----	-----	-----	194
Db	7129	VAGEGVLLAQ	PQGRVRLVELVERGVGVVAVLQRAADAVLRFPLQPGESAHQVGRPG	7188	
QY	195	VNVDSQTIY	FLGL	-----	207
Db	7189	VGVDEHVVRAG	LQGAADADVVRGEQIAVAVADRHRGPGRAHMLDTAVRRVVVDVADL	7248	
QY	208	-----	-----	-----	207
Db	7249	QVEAGVAVOQ	GAETVPVPAQLVGPDPDRHDPRPTLGIHLLRSOGLVVRHFRTARIA	7308	
QY	208	-----	-----	-----	207
Db	7309	SVSWAAGSN	AWSSSTAXGHTAPSGRVRRAVRNCSNAVASTSPNPHIRLASSMTT	7368	
QY	208	-----	TALLRYAQRNCTH	-----	221
Db	7369	SRPVFSADAR	IVSQSTXIVRGSTTSIDASSAMASATARATCTHPVATTVTSSPVWRIR	7428	
QY	222	-----	-----	-----	221
Db	7429	ADPSGTGPS	PSTSTVPISALGAKTATGSGSRSTAESSRXPASEAVAGRVTLMPTCANRL	7488	
QY	222	-----	-----	-----	224
Db	7489	XGWECPAMP	EREPEPGTRXVTGTSTCPVISENLCAWLAI CSNTRXSSEGIWYSTGR	7548	
QY	225	-----	LYNAMS	-----	230
Db	7549	CPAIAAPVAR	LKACSDRGVSTTRPSPKRSCRPLATWVRPIRTSPKTOTRSSRAISSCS	7608	
QY	231	-----	-----	-----	230
Db	7609	AMFSAVMYAS	AALRSSLRVAGSCRRHKGAAPRRGRGPGRGTRTPRPRSACWSRRRWR	7668	
QY	231	-----	-----	-----	230
Db	7669	PSHRPGSCG	CAAGAHATRRPRRRPGSSAPPAPRPSGGRXTGRCGSGRWRDRRPGSVR	7728	
QY	231	-----	RNLFRVPKYNG	-----	242
Db	7729	RPRRPRRRR	PRFPRAPSGRTRPPSPFGRPPAPGPAAGWSRSCSRRRTTAPAAATGP	7788	
QY	243	-----	-----	-----	256
Db	7789	PGXTPAPR	PGXAHRRRRPPPPGRHRAAGTTTPHRRRWGRPPPCRCWRHRRRGXRRRA	7848	
QY	257	-----	-----	-----	256
Db	7849	SSRRPGSGR	RSCTARPASASGLPWRGHRGDCWXRRRRXXAPRRPRPPPGRCR	7908	
QY	257	-----	-----	-----	256
Db	7909	GGCRRSPRT	ATSPGSLRIPXSGAWSGRAGRAVERSSPVSLLPWSSAERPVAHRQRH	7968	
QY	257	-----	-----	-----	256
Db	7969	EBHQOQY	PGDQLDRVLGDPGAARDARTAEQLPAGHRRGGQVPLRDGAQPAHLLRRD	8028	
QY	257	-----	APVKEQFEKK	-----	266
Db	8029	EGVRQHQ	QQRHRRPGGLRGQQQVQVDPDPDERVAPQQQQERRRQRLAGALVPAPADQ	8088	
QY	267	-----	-----	-----	266
Db	8089	QSAAEEDRE	LQDAEHVGEAAQORDARLRHRAEPVGHAPGDGRAGDGHAGDTEHRL	8148	
QY	267	-----	-----	-----	266
Db	8149	SEQPRHOEL	AVTAARHGHAHEVPEDRHREHHAHVEDQPGIAPPVHVGAADHPAVG	8208	
QY	267	-----	AKKTQSTTTP	-----	276
Db	8209	DGPAGPSQ	HGHTGSLRPARLVACRNTSSSVGRSDSTTTPWPCPRSSCTISTRSPTR	8268	
QY	277	-----	-----	-----	276
Db	8269	SRTGASIR	REASASTSGSVPVKBRATXAAADRSSVAARVTIIRSPTRDFRVSGVPSTPR	8328	
QY	277	-----	-----	-----	276
Db	8329	PASITEIN	SASRSRYCVSVSTVPCSTIERTIPTQSLRLGRSPVVGSSRKSRLGTV	8388	
QY	277	-----	-----	-----	276
Db	8389	KNPAARSSL	RRIPPEXCLSLSAASVRSNCSSSSVRRCASLRPTWKRPHISRPSRPVR	8448	
QY	277	-----	YFSYT	-----	290
Db	8449	ISSTAGY	WPVTPMFRRTSAASRVTSRPTARAPPSSGLSVLRMRTAVLPAPLPGGRAHT	8508	
QY	291	-----	-----	-----	290
Db	8509	VPSGTRT	NSPSSAVSDPXVLRSPSATRAAFADPLILLRRSVRTGQRRSXTRSAPADGRR	8568	

QY 291 ----- 290  
 Db 8569 ACSSPGRTRTRVRPARRSFVASPRPGDWRTGRRATARHWFGRGPPRGRRAGPAVP 8628  
 QY 291 ----- 290  
 Db 8629 TGSAAAGRGATGSPRSWHGVRSRRRPTSRAGGCARRPRRRRTTRRRRSXSAGSRR 8688  
 QY 291 ----- 290  
 Db 8689 PGPAPXGWCRRRFRPCRSRSGPAGSRGAXRARRPRSTAAGXGRSRAASPTAP 8748  
 QY 291 ----- 290  
 Db 8749 RPPGTGRSPRTTISGASWPSRSXPGORTSRTAGCCRTGCRGRGTPTSRGRYVYP 8808  
 QY 291 ----- 290  
 Db 8809 DRESPPCRRRTARXCSAGRTSRPGRGSSSRTTRARRPASPAGSAAACWGCXTPRSALXP 8868  
 QY 291 ----- 290  
 Db 8869 HRDCPVRSWARSRQRREXAXRPAPPRIARTCGTSRPHSSGRHRRWRRTGGSAAH 8928  
 QY 291 ----- 290  
 Db 8929 TWDWAGAGSYVPLRLRLRLARLAGQODVVERLGRALDALAGEDPVEVVVDEARAQP 8988  
 QY 291 ----- 290  
 Db 8989 VGQGGPQQPDRTARRPAVLPAVVAQEQOTARSAGELDRSVHGFPGPLGLGRESVGGP 9048  
 QY 291 ----- 290  
 Db 9049 APQVLPVLLPDMRGQHPAARPQOPERLAQMQRVHRVFDLGGEHTEVERAVROAAQP 9108  
 QY 291 ----- 290  
 Db 9109 GLVQLVTRRQVLAALHGLRQPGAGEQVAVRGARSEVEPAPRRPGELLGQRVAEVEQMG 9168  
 QY 291 ----- 290  
 Db 9169 VGPVAGARIGLPEALGDDLLGGLAARARRRRGERAQQAQQVHHRQYSSCGNRPE 9228  
 QY 291 ----- 290  
 Db 9229 IAGDSQSRNYCSTGRYSKSSSRCSPPAARCSRHARSTCCLSTYSSTGTCPSXESTSPGR 9288  
 QY 291 ----- 290  
 Db 9289 YSXXYTLRLCLPTTAGAFCSVRMLCSRASPGRITAGTAPVASHVALSRSAVSVTCSSSQX 9348  
 QY 291 ----- 290  
 Db 9349 LDCGMAQTLSSGAWSEXMPTSMTRPPVTATASEYIWPSCRQASGAPAVLNTMLSQV 9408  
 QY 291 ----- 300  
 Db 9409 GMKEGPTRVSMARSSSGWPASRAITGDLSTYSMRTGAASGWSSVQMVSCMDRVQAS 9468  
 QY 301 ----- 300  
 Db 9469 SIRPAAASCTDSSAGLSRPRSTAPTCRRKAVTCGSRPVSRVSIWHSGTAREAGAVSV 9528  
 QY 301 ----- 309  
 Db 9529 MSNALLGGNGRTGARDRRRSTTACAARRRRPGRGPAPPPASPAGCTACSGRRPAPT 9588  
 QY 310 ----- YRPD 313  
 Db 9589 GRRWCKRSRAGRRRRVRRPRGTAGAAARPPRAGRPPRRPPXPGRGRGRRRPD 9648  
 QY 314 ----- 313

Db 9649 VRCPPRLPGSRRRRTDGGAGRCRGRWRPRGRSCPGQGMWRPRRPAIPXPRGSRGAAVRR 9708  
 QY 314 ----- 313  
 Db 9709 TAAARPRSAGPAGRGACAXGSRRRRHRCPGSRGRRRPVXSRRTGRTSRXAGGRAS 9768  
 QY 314 ----- 313  
 Db 9769 PWPQTDLSSRAGGWSPLPSSSGASASSSSRARSSSRSTARTLSSRCGSGWAPMTG 9828  
 QY 314 ----- 325  
 Db 9829 MTEGWASTQASAWATLCPRSAANSFSRSTAASLAKRSPWKVGEXARQSSGGNTWPR 9888  
 QY 326 ----- 325  
 Db 9889 SAPASTPCASADHPMTPIPAAAQCGSTCFSPRSMMLNGTWATCGVTTEENSAISSIVKL 9948  
 QY 326 ----- 325  
 Db 9949 ETPTYRASPSSTNPASAPAVSAGGTGSGQWTRNTSTXSPERRLSVRLARRKPASLLR 10008  
 QY 326 ----- 325  
 Db 10009 TTPSPRSSRPPPEVMTMSERRVRSRPRPASAPAMPAPXACAVSKKLTFRSNARRTAS 10068  
 QY 326 ----- 325  
 Db 10069 TAGSVRLPONFLSAQVPROTBETSGPCRPVSRYLTQPPVPCAGXGSRRRPGRRATRVPS 10128  
 QY 326 ----- 331  
 Db 10129 ARPPPRACGPRRCRARTXIPXRRPGAPRGSSXRTGTAAWXPRRSPGSGXRGTRAAAPA 10188  
 QY 332 ----- 331  
 Db 10189 GSRAGRPRRRSGKRRXARIIRRRRRRPEVRPGSRGCGGRXWAAPIHRWGSGPCTTG 10248  
 QY 332 ----- 331  
 Db 10249 XRPRTAGCPGAPARGPXPRRCRSRTDRCPGSGTAPRPTGSRARHGRADTRHGADSPRA 10308  
 QY 332 ----- 331  
 Db 10309 APACPRGRPTPRPXVPRRSGASAPCRPVSGLSCSRSTPDVARTVHAVVGHVVRVGAAPP 10368  
 QY 332 ----- 331  
 Db 10369 EQGALLEPGERHLGLPAQLRPGGLDVGAAPAGLVLRRAELGPQLHSRECGDLRLDTDG 10428  
 QY 332 ----- 331  
 Db 10429 HLTGRADVDRAREGRVHQHQSPHRVLDVAORPGLGMAAVDGERQSGERLVDGRHHPV 10488  
 QY 332 ----- 335  
 Db 10489 VAAHPRSGVEQAGNAGVRAQARVGGQORLGEPPFLVVRTRADRVDVAPVALGLRVLG 10548  
 QY 336 ----- 340  
 Db 10549 GIPVDLRGGDHEAGAVRHRQVRLGPGRDPQLRQORQIVQGGRRRQVDDVVDRPL 10608  
 QY 341 ----- 340  
 Db 10609 DGERLGHVGPQOREPLLVGERVEVRDPAGGVVHAGHROTPLDEPPAQVGADEAGTGGDD 10668  
 QY 341 ----- 340  
 Db 10669 GPLSLLOPCHALLRASVALSTTSATSFADSPSPATRRVSSPATAAGEGARSTPSPSFA 10728  
 QY 341 ----- 340

Db	10729	ARRXWNSNTRRSASSTFSAVASRDRVRVAPAGSIRRPQISVSXULARDSCGIPATMPF	10788
QY	341	-----	340
Db	10789	MXQLALPWCTSRQYGRMSSGQLRSTRTPAGSAPSSSTCSGAVSTTSAPCKRTAASMAC	10848
QY	341	-----	340
Db	10849	RSAGTSAPKMPEAELPQTMHRLPRIGPSCHGSAPWAPLXGTXRMGSRGSPGGTATSG	10908
QY	341	-----	340
Db	10909	EGSTAYWIRSRPHSTPYFSKHDSGSAETRSSPGSVSTVPMWPGDTYTLGTTPRSAAATSAPS	10968
QY	341	-----	340
Db	10969	TAMMCTTRSGROWPKLTALSXLCTASIDFPQKSRSPASKLSLSDCIGCPWNGCSGS	11028
QY	341	-----	340
Db	11029	RGPDCGVYMACTXNASRALSIMSGASSTGTGIRPVAATPRSCDGHATRISCPAARSQA	11088
QY	341	-----	340
Db	11089	ARGTMMXCPAOPEAVNIILIRKTPXAVRCLPGARRRSAQAGAGAGAYSMTWVRQL	11148
QY	341	-----NPFCEP--	346
Db	11149	PLAFGEPTWRTSCTTSPWLTGTTGTPVARTFATSASSGRCDPGTASRHLPGPAS	11208
QY	347	-----	346
Db	11209	SQMLRLQTVRGPTLRSRCHGTSAARYGLRTSAXAVGTGPSTNAASCASSRCSRS	11268
QY	347	-----	346
Db	11269	RSQAQCFSKLCTRXTGRSSPSTGKFXSERRSPSVKRVMSRSAPALRAIASGARTG	11328
QY	347	-----	346
Db	11329	PAFAKSRNRNSTAAARSSVAVVMDASSFRLVPAASRAGHAGLAAGRDDVADLCRGPLVQ	11388
QY	347	-----	346
Db	11389	HLDIGEARLLQGAHQRRPVQAHVDAEAAAGLGLGVFADGVDRVPSRSVIPSELPOXWVE	11448
QY	347	-----	346
Db	11449	RXRQSETPMSASSQGRCTWXRSPAQSVASKWPSPANISDTSASRSSTLVTPCTSMTN	11508
QY	347	-----SRNRNAVSEFMKN--	359
Db	11509	SARRPSSRPSRARISTPGCPDIARVSPASSAPSSRTPLAPGNAKXASARSWGGLSRX	11568
QY	360	-----	359
Db	11569	SRVEAVARTPLYTKSFTHGSSSTRRRISRSFTVALPPLSGALPSETAWAGDHRSAPLS	11628
QY	360	-----TH--	361
Db	11629	RSATHCAHSGTRSPPIISKEERSAGTEAAHAAVRASPSAGVSAQSPKTSRASSVAV	11688
QY	362	-----VLIRNE--	367
Db	11689	XATITSRAGRCMGLWGTFAAARTAWASSSTVLVRDEPTLKARPOAFSVSGARTWTSTM	11748
QY	368	-----TPYTI--	372
Db	11749	SLTXLNARTXPPSPXTEIGSPRIWLKWMATALXGSRMFCHSPYTLICIRNAVKGPPWR	11808
QY	373	-----	372
Db	11809	IVSSSRSTVSFALPXLISAGTIDDSRIGVSPCYTATEDAKQKNETPVRSAAALMRFTLP	11868
QY	373	-----	372
Db	11869	ITLVFXLRLFTEKLIASAANAANKWTWNSRCSAKSLSTSVTDVPNRCAPSGTRSAPV	11928
QY	373	-----	372
Db	11929	PLRSSRTDTRCPLRSASATCEPMKPAPVPTVQCAIKSPGSKGKQRETKNSVSEGEARGR	11988
QY	373	-----	372
Db	11989	ALPXSUWLSARPEPCERTABARSTAPSPGSRRTSARCERGLARTTHGVRSRARRGRRT	12048
QY	373	-----	372
Db	12049	ASSAGRVPSGRVHARPGPSRRPGRPVAVASVPDWSARSPRCRPPAPAAARVRRPAGCPAR	12108
QY	373	-----	372
Db	12109	GRTRAPSPGCRHRPRTAGCCGRPAAGWXTAERPSTDRASCRTASRSPAAGCGRRR	12168
QY	373	-----	372
Db	12169	GRPCRRPWTTPRASWGRSARPPRPSRCCRTPRRWRPARSWPCRPCTPAXVPCRPAPH	12228
QY	373	-----	372
Db	12229	WYVXTAPRRSTGRRAIRPRSPDPDRGGRPPRRAAWHRRRCRRARCTGRRPAPGSPGSR	12288
QY	373	-----	372
Db	12289	RSACPARSPAPDRRFPVRAARTGSRXSLPHRXPGRAVPVPHPCRLACSRKPPSPSRPPP	12348
QY	373	-----	372
Db	12349	POACSTADLDPAVVSQHPVRCRRPGGVHLGLPADHRVLDVTGDSLDRSATQHDRVLDLA	12408
QY	373	-----	372
Db	12409	VRDPAVALDGGEGTEVGLHHRVRADDDGTDGAGDLRPLRLHDSAHQFAGLVHLAVDA	12468
QY	373	-----	372
Db	12469	GLQGLQQVPHLQVGDARDVLVPVDHHRGADLVVPVVGQLOGVGDQLVPPGPDAGHG	12528
QY	373	-----	372
Db	12529	LVHGRAEGVHADDRQVAPCVLGLLHHRGHPVVLVKEGHAEPARVGDGEGHDLRGGLVAAB	12588
QY	373	-----YGTLL	376
Db	12589	VRHEVDDAVHDVVVAQVHDEVVVAEBVTGHEGVGPARGVLLDVGDVQAEALRAVSHGSL	12648
QY	377	DMS--	379
Db	12649	DLRRGRPDNDPLRYSLGLDGIQYIGENSSIGHRIDQLFGGRMNRPPQPRANPAHQYERLH	12708
QY	380	-----	379
Db	12709	IPHLAVSVSRGFRKGTPTDGIETLPTTIVLACFSGFPXVSGPAGAAALSPPRPLLTXYVTE	12768
QY	380	-----	379
Db	12769	SVRSGMAVRVRVFRFAEQVDDVADGGQLPEDVGRHVPAEALLQPAADHRRHHRVEPL	12828
QY	380	-----	379
Db	12829	LHORAANVQGRGESHLLRLDLHLVHEHPTQLRRQTRERHHVVRVVAARLVGVRGRL	12888
QY	380	-----	379
Db	12889	LDHRVAQAAQARHRSALPVRDQGVHGVGDAGIERDLDRRRLRWRAVDGDLVDVHDHAELR	12948

QY 380 -----SIXYNET----- 386  
Db 12949 EAHEGRRDRHHGVQGRGVRGRLSHVSQYVETETGGQSLDAPETGRGYAHLGPPGQD 13008  
QY 387 ----- 386  
Db 13009 AGVGRPEPVLARVQQRPRQTEVRRGRPGGLGMDPEPGVVRGVRDRPGGQPHILALADD 13068  
QY 387 ----- 386  
Db 13069 LDLPLTRQHRVHRHVLGEAALFALGAGRQERDRQHGGDDLHHA VEGAARVAARVAVVQ 13128  
QY 387 ----- 386  
Db 13129 RAVPQPOVAARDAAGRAQQRVGVDVVAALGRSRPGDGTQRAVLPGVGRQDFEVPVRLER 13188  
QY 387 ----- 386  
Db 13189 RPVHAHAVAGGAHGAEGFGLGVAPXRSDHVLHAAFQORLVDRVRDHGVRGQOLDER 13248  
QY 387 ----- 386  
Db 13249 PNAVLGRRDRGGEPRHVAQVRPVRVAVGHRGFRVRRERRVRCAGTQLVGGRAQP 13308  
QY 387 ----- 386  
Db 13309 VEQGVHRAVEGHVGGDPYDQPLVGPAADLLDLAGVTGDHGRRRGRDRGRHVIARE 13368  
QY 387 ----- 386  
Db 13369 PLGGLVEGTDHRRATPRGLAQORPTADQPGVGHGQRPGGDGGDLTDRVADGGGL 13428  
QY 387 -----MFVEN----- 391  
Db 13429 DAPXPQGEPELHREDHGLGAVDLGVVALQOHLPGREPDLLLEHRLQFVDHLGEDRLGF 13488  
QY 392 ----- 391  
Db 13489 QQFAAHPGLGPTVREHPDGTAGIRQGPQDETGRLSGQRAQPGQFVPAGRHGRPV 13548  
QY 392 ----- 391  
Db 13549 DQRLGFWGQVPDRGQVSALLDPLGERGRAVTERLRGGRHGRQSGRLRRGAGRGRL 13608  
QY 392 ----- 391  
Db 13609 FEDHMGHGAVAETGHSRPAGVRTLGRQPGRQPDVEAGPVDLGVRAVQQAGRDPVLV 13668  
QY 392 ----- 391  
Db 13669 EGEDRLDDPGDPCGPPLQMPDVGLHRPDRARLLTGAVTSVDGGDRADLHGVAEAGTGAVGL 13728  
QY 392 ----- 391  
Db 13729 DEVNGVSDARGQAGPHDVLGLPAGGHDVAGAPVLVERAAAHQHRFPVTVAYGRQOL 13788  
QY 392 ----- 391  
Db 13789 DDQHAALAAAVAVRRNGVAGSRGRERLAPAVGRQAGRAGODRLARXHQQVDPAGOR 13848  
QY 392 ----- 391  
Db 13849 HGALAEPOALRGVQRHOGGRARGVDGEAGPAQVERVRDVAVGHARRADPRPCVDVRQV 13908  
QY 392 ----- 391  
Db 13909 FQPSGAIWVADPCEHTGVPCQVRDARVLQCLPGLHQALLRVLHQGLARRDAEEL 13968  
QY 392 ----- 391  
Db 13969 GVETRYVVVQETAGANRRPORRIEPRTEVPVLPFVAVGGHLADGVASGRQQVPFAGRVDT 14028  
QY 392 ----- 391

Db 14029 TGEPASDPDHGRRREGVRAIKGYGHIPFVVVAGIHSSQSGDSRTKXTALDTRPGAVA 14088  
QY 392 ----- 391  
Db 14089 SCRTWSPGASSFVRTMCSMDSRWSLAPOVADLPAEFPPGQQRHPPGRHSALLDHFAXXF 14148  
QY 392 ----- 391  
Db 14149 MKETPATSATPMPAPARARSATSIWSXISAIATKAGLRVANCWPIRALKSVARPRVMLG 14208  
QY 392 ----- 391  
Db 14209 GNNGSSPRCSDISGSPLEATEQSRAPKRLQHOGGSASPPMKPGPIRHSISGVPXLSA 14268  
QY 392 -----KTASDSNKTT----- 402  
Db 14269 ATISALVFGSACRXPJTASSSALNTPLOCLISRAMSPPTRCRYAWGGQPSGWCQV 14328  
QY 403 -----TSPSM----- 407  
Db 14329 LAYSSRSNSVGDPPHLFTACCAARKAISGGTSPAMATPPSSCASASAGXKRPRSTAHST 14388  
QY 408 ----- 407  
Db 14389 LTHGGTKTPRPMYRSAICTNPPPCFQAYWTRRARCILLTIIVSVLSVFLRESRPAPA 14448  
QY 408 ----- 407  
Db 14449 FPGQNSPHRRORTTELAFVRAGAVRASRSRSGSAGVMTVAHRVTMTSMVNNAGVRTPS 14508  
QY 408 ----- 407  
Db 14509 SSPWRIITSSVSPQFPMSTPMERRASTRDTISAAAAALPASLPTTATSRATPAHTRAPP 14568  
QY 408 ----- 407  
Db 14569 PRPSCRSVFRPTEAKNGNSTAVTSSCSLPRASARAGAVFGQMAPNRPKAPORWMPSS 14628  
QY 408 ----- 407  
Db 14629 VATADSRTPGMISVSTPPSGSRRSRTRRKAGRTSTMTPKKTAVSARVSTPRGEP 14688  
QY 408 ----- 407  
Db 14689 RNRAIRPASSQQARSAAVAAPTMAXVPARDWSIPRSISSRASTGKAVTLMAMPKRKVNR 14748  
QY 408 ----- 407  
Db 14749 KPTSGAXFVYSAYPRAQFRANDTQTPTIAPVRPAPSMWVSPNSMPIRNMNSASPIW 14808  
QY 408 ----- 407  
Db 14809 ATRSRXEWRAKGRASWTFGATRPSTEGPRTIPEIISPTTRGWPKCRPTSAPQARATAT 14868  
QY 408 ----- 407  
Db 14869 TARASSNRHASTCGASPGVALPEPGRDASSLPVWPCHSRAAASGSDSRHTASTRVVP 14928  
QY 408 -----CFQRTFIDP----- 416  
Db 14929 RDVTSATRSHNFQGRQKQAAASGGQEVAFPGDHRVPVDPREDQELVGLKRPGLLIGDDR 14988  
QY 417 ----- 416  
Db 14989 DVRACRVAELVLVHLGDAGQERLVQAAVLQDDVALRGA VAKGVQPLVPRGGEERFQESV 15048  
QY 417 ----- 416  
Db 15049 TGLLGPGBELRVQVRLSGADQORRLVDLARGAAAGQHEREASAVDGGQLGIHHLQAGA 15108  
QY 417 ----- 416

Db	15109	LGHPDLAQPVVQVLVADHVVGAGVEHRCQIALLGHPDAVVRQQAADLPHELHGTQVV	15168
Qy	417	-----	416
Db	15169	EHRDRGDLGAAPAEQVVCBEVGDQLDARRVGRGEDGGGVHSDQAGPVGVRQOQAVV	15228
Qy	417	-----	416
Db	15229	AADVHQVAGAEVDQAGDGDGLAAQVVRHGAVEAGPVAVALAVEAGVGAQLHOFATA	15288
Qy	417	-----	416
Db	15289	GRAAHQLQRNVHVLAAQPREHAGEGLRSEVEHQQLGVVADPATVDDGRVTGRAHDSPV	15348
Qy	417	-----	416
Db	15349	SARLRLSQPMTSAASQSGSLTRGNWSSLRAPAKTEREPTASLISRLLTVVRAPTSHSP	15408
Qy	417	-----LM-----	418
Db	15409	SASRWEATQPPHSTGAKLRDSPFCAXTCGCSRMSMTPLVLLSPYMGPRRTTRWPAQSS	15468
Qy	419	-----	418
Db	15469	SASARFPXGVSSWVRBAEMVLIWTKRSHSQRSARVLSTVSRSSSGCCRAQXRREGQA	15528
Qy	419	-----	418
Db	15529	ALTSTSGPLRRASSXLARVTGMAEPSARVAVPTNPPAPTATLTGAPACAAAARPEP	15588
Qy	419	-----	418
Db	15589	AADRIIPRAWXWXTSGTRWGRSAXPRPASWCRPRARGXIPLCWGLRRRPMXSARP	15648
Qy	419	-----	418
Db	15649	SPRRTAGTROSAGDNAGRAPFPGTGAAYXIPGWAPCGPRKXFRARPAPYPGYRSAAAHR	15708
Qy	419	-----	418
Db	15709	FSCELHSHFATDSQASRIIBARVRKFLVLSAAPMIRVPSADNTAXRPAPFQAMPVTRG	15768
Qy	419	-----	418
Db	15769	FAAGTGTQGRMPLHPGAPQLPSPRRSLRTAMIVKSGQAPGRPNPHLLEQNPQVEPR	15828
Qy	419	-----	418
Db	15829	PAPERKSTTAGERVLTGKESAAGAWSRPGALDTSQHSPPREYQVTRAPTTLSSLDAKIIE	15888
Qy	419	-----	418
Db	15889	RIAAGERTTICASKLYLSRQVEYHVYKMMERFRVFNRTALVAKAYALGILRSEVWPRV	15948
Qy	419	-----	418
Db	15949	LOBRVNKATVTGSGRLRRAGQNAHVANAKVYRNGFSYIPQRSVDYGMAGRPKQVTR	16008
Qy	419	-----	418
Db	16009	PVSGQARRDAGPFEVHRRRLGAVAPGGPVAAFLPAXLPCTHARAALSALPTPRALLT	16068
Qy	419	-----DYLDLSL---	425
Db	16069	TAGGTAPQGRCDKMIKMSNVQDPFVAVVVSFRLHSSVAPFGALILDASDDLGHHA	16128
Qy	426	-----FLDEI-----	430
Db	16129	RTTRSYGGOTLPYSGCSDSDGDRFLHEDQFPWLEGLSQVHLLSDRELQVFLLSHGYS	16188
Qy	431	-----	430
Db	16189	NREIAALLGVTERTVKAHMAQILAKLGVESRLOAGLVAFAHGLPGHIGDAEGSDHPPDTS	16248
Qy	431	-----	430
Db	16249	AGSXIERSALRGESGQDRTRSGTSTSPRWKAAASRCPSSLKCSGSGVYRGSTYRMAVSG	16308
Qy	431	-----	430
Db	16309	PARLGPMTXWTFSSPSPGARSRTTSRVCSRTSVKDRXSPSPSGSCQEGEXMWDAMTSGVEGC	16368
Qy	431	-----	430
Db	16369	RSMIRTSPTFRGXVNWFRAPXPMFPMTHVIRSAGRSRVAKSGWRXQAPATASEV	16428
Qy	431	-----	430
Db	16429	RSPSSVATVAVQVRVGDSEAPLPCDSEAPXWQRGAVVSRPTSAPAGNSTPCKTNWFRP	16488
Qy	431	-----	430
Db	16489	GHSSTRSLPPGSAFTRHVWTSRVTRGLATGLPSTRTSTCAVSGGPGISGAGRAAYR	16548
Qy	431	-----	430
Db	16549	NASPSASVRKVTETPVGIAARPTPPSGKNCRSXKIGROGAIIPPSTSWRETAPRXVSTGS	16608
Qy	431	-----	430
Db	16609	FARMRRPASNIPIHLLSSRSGRPTLRDXQRCGRSRRFFSTFLRKQFVRXMTTDTRECPRRR	16668
Qy	431	-----	430
Db	16669	NAAILEXWLTKRLRTGTPPGVQWVWWSVPSVWRTSAGICGTSTSTXTVTNTSCCWGTR	16728
Qy	431	-----	430
Db	16729	XHLRPRRSPRTRPGRCGSGWPGREGRRPVTRCCNGPTGCCSWTTTCGRWPPWP	16788
Qy	431	-----RNFSLR-----	436
Db	16789	SAPRSAPGRTWVRCGRDRESSAGCPGSCPGSSTSSAGTATAGTSTRSGAPAWASTPSR	16848
Qy	437	-----	436
Db	16849	AARASPWTWXAAPASTCSPVRTPPSSRXRRRRVRASCLLPEHAVHVDVDPGLRLRYLL	16908
Qy	437	-----	436
Db	16909	RRAYWGAFFGATAPVGGGTPOGADPAPHRAGVALHTTAVLRVGRHRRRRRPTPAAPA	16968
Qy	437	-----	436
Db	16969	QVSHACCRTEPERTQAXVAAMPFSSGTDVQVPSAATRVTSRVLCWISPRRGAKRGGR	17028
Qy	437	-----SPTTV-----	441
Db	17029	SRPEAAAMWSRTSTVLPVSDVPDAGTPYVGNATNAATSTPTXSRVCEPSPNTVGT	17088
Qy	442	-----	441
Db	17089	PFSEASQXWAMTPASPGLCRGPYTLPSRNTLCSVPCSLRNPMTYFSAQYLAVPXGAPG	17148
Qy	442	-----	441
Db	17149	LGSVSSFVGTGTPSPKXSAAPPVEANTTAPTACAASSTLIVPSLTLTRASTAGVATETRTSI	17208
Qy	442	-----	441
Db	17209	WAARWMTSGRADRTASSRAAAASVTSTRMTARGSRSSGSPVVRSSSTVTTSAPASRKAST	17268
Qy	442	-----	441
Db	17269	TCEPMKPAVPVTSRRVPAENPGNMDMARPSOLRHRSAAIVGRGTGKTSIWAIVAPTAA	17328

```
QY 442 ----- 441
Db 17329 RLHLGRPTAEAPVTCVSGSMELPRPGPRHYDRFEATEAVPKYWGVDXRXYRTATPWG 17388
QY 442 -----NLTPE----- 447
Db 17389 STVPISPAVSCSTPRRTVTPSPSAMRSPRSSGSPASPATVRXRCGTRAAAVTSCXWG 17448
QY 448 ----- 447
Db 17449 CCTGARGCXSPMWTTPRWSSRRRISPCSRPGXPRASNGVSRASVSPPIWRPRP 17508
QY 448 -----HRRVN----- 453
Db 17509 PAGCSAXRRRAVCRAPASPTSSIRVRCPPSPGPRMWCWPTCTCGAHLGRAGARA 17568
QY 454 ----- 453
Db 17569 AGGPGCCARSRRRCPRMXSPSPATAVRGRWLPXRPWGXRSAPARYWSGQPTCWRQAP 17628
QY 454 ----- 453
Db 17629 EQSRNLNFVFSVFTSVSPVPFPXTHSAGCYRCVLGRAEWRKPGCSXPGIFIGRDX 17688
QY 454 -----LST 456
Db 17689 SARPLPMWRRRVKRAVRSGSAARFRGWTSSRGIFRWSRTCAGEHILCSCSRPGSICRT 17748
QY 457 SNSLW----- 461
Db 17749 SRSNWPARFPGTGAXXSISDGLHGVATDSSVGTSHIENMRRLYSALSDLVLPKLTG 17808
QY 462 ----- 461
Db 17809 PLPDGAEEFFCFGMPLVHRHRELGPEDYELQYIGSNWGRWKPFIETVEAARQARLRXI 17868
QY 462 ----- 461
Db 17869 RVCGRWMDGETCPGFEDITSAGAGTAEAAASRRRCSDMSSPRWDGSSRRRCWSARSWP 17928
QY 462 ----- 461
Db 17929 TRGCSRPGCSKRRRAPCRCCPRKNSCPPCTRVTSLRXWARTPLPRWAVMOATSSGTR 17988
QY 462 ----- 461
Db 17989 RSWGSGRTGRGPGTTTPGSGSCGGSADPLPYRFVEDRGRGVRRVQRARGAGHRDADV 18048
QY 462 ----- 461
Db 18049 VAQRTPGVAEPGLIADQQQQLGEVEVEHIAVALVVRTDQDRRAGGPAGLHPVHHLAR 18108
QY 462 ----- 461
Db 18109 XRHPDGGQGEQAGARDFRFRVVDVGVGRPHQDPVRAQIRRANDAADVPTGRPVEQVA 18168
QY 462 ----- 461
Db 18169 QEVLPDRDPVEVVGHLHDGDDLGRGLVLLPOLHQQVGRHRDLLRPHRTKLLPRPLRQLP 18228
QY 462 ----- 461
Db 18229 LSVVEERAYGPAVLHRRERADSHHQELTGARALVAVEOCLPILLETSVAXGDPSPRRRP 18288
QY 462 ----- 461
Db 18289 LSTXSCHSKBPIFRSMHLONDCAPRLEAGQAAVTSHGINAPCAQHGRYVVRPPEDEPL 18348
QY 462 -----WWL 464
Db 18349 DPAWRPVDLTRAPFEDRAEDRAPWPDPRFALCWNL 18384
```

```
RESULT 13
ABP76679
ID ABP76679 standard; protein; 19938 AA.
XX
AC ABP76679;
XX
DT 26-FEB-2003 (first entry)
XX
DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 3.
XX
KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
XX medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX
OS Streptomyces viridochromogenes.
XX
PN WO200268436-A1.
XX
PD 06-SEP-2002.
XX
PF 24-AUG-2001; 2001WO-EP009815.
XX
PR 25-FEB-2001; 2001DE-01009166.
XX
PA (COMB-) COMBINATURE BIOPHARM AG.
XX
PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX
DR WPI: 2003-018650/01.
XX N-PSDB; AB237515.
XX
PT New avilamycin derivatives, useful for treatment of infections, and
XX nucleic acid encoding avilamycin synthesis enzymes.
XX
PS Example 1; Page 68-301; 319pp; German.
XX
CC The invention relates to avilamycin derivatives (I) with antibacterial,
CC virucide, protozoacide and fungicide activity. (I) are useful for
CC treatment of infections (bacterial, viral, protozoal or fungal), in human
CC or veterinary medicine, particularly where caused by Staphylococcus
CC aureus. (I) are more hydrophilic than known avilamycins. The present
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-
CC AB237516)
XX
SQ Sequence 19938 AA;
Query Match 21.8%; Score 534; DB 6; Length 19938;
Best Local Similarity 1.2%; Pred.No. 0.013;
Matches 231; Conservative 105; Mismatches 126; Indels 18744; Gaps 79;
QY 1 MGRK-----EMVVRD----- 11
Db 615 LGNKLNIRESDAICQVIRAFKDNVHVHVGKSPKDDIETINELLADLQTIKVL 674
QY 12 -----PK-----MFVLISIFLL 24
Db 675 RLQKDSRIKKLAPKPKVAVEAKEILRGDTLFAHGIVQGTERNELLDHLLTTKPF 734
QY 25 V-----SPINCKV----- 32
Db 735 VFNVDELDIDEFKNEQALVAPAEAFINAKLEADLAELDDDEALELLOSVGQDEPGL 794
QY 33 ----- 32
Db 795 ATLARVGFNTLGLQVLTAGPKESRAWTIKKGATAPEAAAGVIHTDPQGFKAIEVIF 854
QY 33 ----- 32
Db 855 LVETGSVEARAKGARMGKDYVMQGDVBEFRFNVAQFHDAAFKAGKAGKSTPPGG 914
QY 33 ----- 32
Db 915 GPLLPVLCWCSXSSISLCPAGGQEWYVIRYHLREGVESCRXTASEARLSPRXSRR 974
```



QY	33	-----	32	QY	62	-----	61
Db	975	STTITRSPSPSTATPYSSRPRIMORCAPARTCCALRRTPVDCSRRTPLPTSMCRSG	1034	Db	2055	SKPGQVSPSHHRPQTRIRRSRACLAASSTTMNGFHRPQLLPMYCSXSRSRSPSGRGRTS	2114
QY	33	-----	32	QY	62	-----	61
Db	1035	SXSIRIRTLVRPREACLRGSLGLXHLVLAQERPODARPHQQAHRGQARPLHGDRQTR	1094	Db	2115	WGIPKQKNSAPSGKGVNLCGCTRSLSAEXSRRQFSIEWVTELSAVATPKWPFSEID	2174
QY	33	-----	32	QY	62	-----	61
Db	1095	AEVPLAGGVAAADRRRTTPRVPGYGGDRDPRCPVPLLTAAQVVLPWMDGAGMPSNPMF	1154	Db	2175	YQAPVPGKAGQDOLLVRQILPGLHEHKMCSFAHVLHRKMPDOLLVHPRKRAADPDLT	2234
QY	33	-----	36	QY	62	-----	ILR 64
Db	1155	POVGALXLMFRFNVAQIHHVLLNAQVRGLVIGPLALSRASGLGVAOSSPADRAV	1214	Db	2235	AEFTRLHTGGNGLAESRLPMKIPGYEHPGRHSALPRTHLXPPPLALXVXGKTGTTLVK	2294
QY	37	-----	36	QY	65	QLETTISTKYNSK-----	84
Db	1215	GPCPERQRPVGRDHLAVRTGEDLGVGVAEAGQGEAVGAAGAEIAGRGQVADAFBERG	1274	Db	2295	TLEKTLGTFKFSRDRCSGACLOHVGCPCDOYGRAGADLPLOGLYGSHRDLTAAVGDGHRM	2354
QY	37	-----	36	QY	85	-----	84
Db	1275	ASAYADQPGGEAXAPLVAEEGLPLMGDRGCVVDEVTEAFVGPPGPDQDATAFOVANA	1334	Db	2355	RGORRRRAQOGPPPAARAPALPSACAPVRQVQHDIGRPGEDGGQRTRIEDVRGADGA	2414
QY	37	-----	36	QY	85	-----	84
Db	1335	RAVRILARHVALDHVAALLGLVLRHRPPLRVQIGTVDPGGDEDAADGGFGGQGP	1394	Db	2415	RORTALRQALPPAGGLRGLQIGLAETLALLTPPELARGQPRREOGEILRRELQRG	2474
QY	37	-----	36	QY	85	-----	84
Db	1395	GVCLPDQVFPGRAPVFPVQAGAGDAVRAGVHPPFPGVEDHLGVVGVVDDLGEVGGRR	1454	Db	2475	VHIGDHLPORAPVQOPQHRQEBVTAAARVPORHRTVAGEAGEGPLEDLGRQADGGVTT	2534
QY	37	-----	36	QY	85	-----	84
Db	1455	VRSGPRAHSITLGSASSRAIRAGQQRGSATSKTSMVSRPKLPRVGSPPMIVRIASRP	1514	Db	2535	RRGEHDTAGEIGTVDPPLHGVAVRTHLSTPOYFGTASVAALKRSXCRGPGRSSQLPDDT	2594
QY	37	-----	36	QY	85	-----	84
Db	1515	MSTSCARAAEASSPSCWTRVSMGVVVRVAMTPEVGHQSSAGRSSGHGARSSAAR	1574	Db	2595	QVTGASGAVGRPRXSRAGVAVTAQISEVPNPVRPTIAALRCRSCBEGRAMSMPGFSAG	2654
QY	37	-----	41	QY	85	-----	84
Db	1575	SSKGARVSTGRQAGSSGSSAGGRTYRRPCWAQGALIPCDVTAAWPASRRGAQSFCCK	1634	Db	2655	TRLLVTGGAGFTGSHVVDAPLEAGAEVTVLDDLTGDPERLDPRAVIRRVDTVDAALDE	2714
QY	42	-----	41	QY	85	-----	87
Db	1635	QERKIGSFEWQDHVESGRRLGSPHATLVSSGRHCSATATSARAPVSSWXWESARSRS	1694	Db	2715	AVRSARPDVICHLLAAQIDVRVSVATPAVDARVNVVEGTINVLEAAHAHAGVAVVFASTGAL	2774
QY	42	-----	41	QY	88	-----	89
Db	1695	RWSTAGPYARSTTDSGCRSGRSPFVRCGRSRWRPRTCXSMGRRTRPRRSPSPWR	1754	Db	2775	YGEVVPVTNEDTLPRPGAPYGTAKYCAEKYIGLFNRLHGTSHSVLRGLGNVYGPQSPG	2834
QY	42	-----	45	QY	90	-----	89
Db	1755	CPPTTSGRSGRTSWAYCSTGRVGRGTSAAFPARMMPWARTGXSWRGTPPTTTRNRSG	1814	Db	2835	EAGVTAIYGLASEGGVTVFGDGSQTRDYVTVGVDAAFVAPYGTVPASGTSDTGKS	2894
QY	46	-----	45	QY	90	-----	89
Db	1815	RAPARCSPPGCRHARXWTGWRPAGPPARRSSWSVTSTATMCSTSTSPSRPCCXS	1874	Db	2895	TVLEVLDHIAASGRDLPFRPAPRPRGEIQHSTLDVTRVAADLGWTASVPLEKIAATYA	2954
QY	46	-----	45	QY	90	-----	89
Db	1875	AMRQPGSATPGVRCATTRSASRWPAARXTRTPRPSSTKRYGSGSAEBEQQLPEQPG	1934	Db	2955	WVRSGSPVQOAXLTCAGAAAGVGRRRSRWRPRTRRTAVVCSATPARCAGSAPCGVPPP	3014
QY	46	-----	48	QY	90	-----	89
Db	1935	VVVGPRPVLDPDHLRPLEVACQTAQRGSGVRAQRQSLDVTLVHGGHEFRFGHQHQG	1994	Db	3015	TGAVAPPNAPCOXARRRRXRRVSRPGSTWTTACSGRQLARTRLLHQRDEGGVLTGLQV	3074
QY	49	-----	61	QY	90	-----	89
Db	1995	ARRERFHPGREQPRVGERADQHRDDGFSHLGDDMSRHRRLDLDAASAVPAEAVMS	2054	Db	3075	DAGAAQVQVEGEALAAAREVGVDHAGAPDLDPVAVAVPADEVELPRGQEPQPAELSRRS	3134
QY	49	-----	61	QY	90	-----	93

Db 3135 QPRTDHRPGBERGADGVHGGGLQPHVVVHEQHVPVGPFPQHRVTVGRRQPSRGPODHPRP 3194  
QY 94 -----  
Db 3195 PGLVHLGGDLGLGRCHRPVQQHDVFTVQLVEVQMPADVLTQGTDRHHHTWTPPGGVP 3254  
QY 94 -----  
Db 3255 VRSLFLVHSHSSMAAFRRRGHSRVS VVHRTNCFLEFLNVLKNRRLRPHRCKRSRSGVLP 3314  
QY 94 -----  
Db 3315 LDRXGMLEAGRLIRANDPVLTYRGAVSLQEVDDGGIAPWRIPFQERHLFPFEGGVGRAAMP 3374  
QY 94 -----  
Db 3375 TGVRTFRTDAEGLAFRYAARPAPEMPGPETAHVDRVDRGKPVASLPLVTDREVHTCRV 3434  
QY 94 -----YILAG-----P 99  
Db 3435 GALPGGSDRLVELWLPCLNQFVLHGVLPAGAEVGRDTHTAPRWVHYGASESQRGALSP 3494  
QY 100 IQNYSIT-----YL-----108  
Db 3495 TRNWTAIVATELGLDLTSLAIGAGCYLOPLFATLLRDLPADLITCWGMNIGARALNQF 3554  
QY 109 -----WFD-----111  
Db 3555 TYRPNLVGLRIIRERHPSTPLVIASHHYSPMDPLEG DYLSLTVREQTREVDLLRA 3614  
QY 112 -----111  
Db 3615 DGDNVHYVHPSLAGPETAHLVYEPRYTDPLHFNQEGHDLAAAFQKVELVPLDVR 3674  
QY 112 -----111  
Db 3675 XLPDSPKALRIKEPADVSGGWS DPSAS PMCPGSPMAKATS PAXRRDSTPSPASIWAIXA 3734  
QY 112 -----111  
Db 3735 LTVRSVTPSSAAISRLEXPMLSRKKTWSRSESKCTWLSPSQGNSSWNRSPSLES 3794  
QY 112 -----111  
Db 3795 QPEYKVPWPYERVVLACPKRSSDASSI KAPKATTELXREGGDTHATNGSWTLILII 3854  
QY 112 -----111  
Db 3855 LSHLRPGWGLVPPAVVSRARRGVGKADKAARARCVHGNHAGRNAATGPPGATAPRRXC 3914  
QY 112 -----111  
Db 3915 TSKNPASRRPACPETGRVTCFGRPAMRAPXSTDRGIYENFPFRATFALATCAFCPARRR 3974  
QY 112 -----FYSTQUR--118  
Db 3975 RPLPPTVAQLTRSCRTLGHTSDLRIPSAAXALATRAVLGTRKRSIILLTWYSTPWRLR 4034  
QY 119 -----118  
Db 4035 XSLLAMQVVRSPAMRSMILASRLDKVVGAVLTWYSLGEBCEKESVRAPGRDQAPADS FV 4094  
QY 119 -----118  
Db 4095 PNTRSPAVVDFRSGAGRSWTLGCSKWKGGFRGPAAGXPDFTIMAVKRRLRLGSGXGAPG 4154  
QY 119 -----118  
Db 4155 CNGRIPRCGVPPAAKPRVTGIWPAGKLAGRHAVLSADGTRIMGADTRNFLTRARLIL 4214  
QY 119 -----118

Db 4215 DACESVAMNWRHRHNRCAAADLXPGYGAGRARRNPHLPGPAHPCIQYAAVPFGKGARP 4274  
QY 119 -----118  
Db 4275 ALSPALCRVPANVRGLGRADHQRRRRPQOLQGIHPLRLARGLHQDAGRGHALRPHLVPD 4334  
QY 119 -----118  
Db 4335 VHRHAQALGRMSAAGSGRAAAAHAGAPVRVAVVGAGGFGVTAVTRALGSAAPVTRAN 4394  
QY 119 -----118  
Db 4395 YEEARRSGPPDFVLNAAACPSRRYWARQHPDDDRRTVDKTRALLRDWENDRFVQISTISA 4454  
QY 119 -----118  
Db 4455 RTQLDTPYGRNRAEABELCAGHLVVRGLGPMYGDNTKGVLDILEQPVVYAHGESRQSPA 4514  
QY 119 -----118  
Db 4515 PVWCGGWVASHLDAEGLWEVGARTTVSLREIRDAVGSRSVFAGARKDDQFPLVSEPDWP 4574  
QY 119 -----118  
Db 4575 DAADVIGHLRARSALTGESXARPVTRPSSTVAGSATT PSCCPCSTDRRSPACSRGPA 4634  
QY 119 -----118  
Db 4635 ARTCKTFRWSMAARPAVANWCSCATPPTGASCTARATATGPASTAPWRTTCAARSPSP 4694  
QY 119 -----118  
Db 4695 AMSTAPATWSWTSAATAPCWRPTPTGPAWSEWTPPPSSPRPTRRASSXSPTSHTT 4754  
QY 119 -----118  
Db 4755 CSAGAAPRSSPRSCSITTCRVWSSCGRSAACXRTTASGSPRAICPRCSTPAPTWSAT 4814  
QY 119 -----118  
Db 4815 STWTTTGCARSSGWPSAPAXRWMPSPPTABASRSCWPAAPLARTSRRWSPASAPER 4874  
QY 119 -----118  
Db 4875 RTCPTRSFGGPRSPVTDWSSNPPRGTRGCTPLATAPRRATSSCSTAATRRSCPASP 4934  
QY 119 -----KPAK-----122  
Db 4935 RXTRTSSAATRDARTSRSPRRRPGRLSPTSWSRSGSTGTRWSPGNATSWPPEAAMSFR 4994  
QY 123 -----122  
Db 4995 CPRWKLCDRVAEVTSGTRTVLAVCLLSVLPLLAALLLWHGTTCKDDASRPGSGSATPG 5054  
QY 123 -----122  
Db 5055 DAPHVEAWELLALAVVAVARACGALVGRHLGQPRVVGEMISGIVLGSVLGLVAPGVH 5114  
QY 123 -----122  
Db 5115 DALFFPAALHSYLDLVAQIGLALFMFLIGMEFGDTHHEGAGRTGAAGVIGVCVSFALGCA 5174  
QY 123 ---YVYSQY-----128  
Db 5175 LGYALTYGAPDGVGLPFTFLGFIAMSVTAPFVLARLLMERGMLQSRAGTYAIVGAATA 5234  
QY 129 -----128  
Db 5235 DLACWLLLAGLVALLRGSPGLGLVRLTALTAVFFGVVWVVRPALRRVLPERRLPDGG 5294  
QY 129 -----128  
Db 5295 VLTLLIIPGVLLSAVATELIGIHILIFGALFGAICPKTAPALADARKLQELVTAVLLPPF 5354

QY	129	-----	128	QY	148	-----	147
Db	5355	FASVGLKTDLLQGRGGALWVWAGVALLVAVVVGKLAGSAAALMSVERVDALRIGVLM	5414	Db	6435	SRVTGPGGWAANCWKPSSRSPRXSTNWSRSCRRSSGSRPGCCWRATTPTSTAPRXS	6494
QY	129	-----	128	QY	148	-----	147
Db	5415	NCRGLTELVLITIGLBELVLTALFTMLVIVTLCATVMTAPLDDLDRABARTAPARTKA	5474	Db	6495	SRCSGSPRCGGHAGSSPPPSATRSVRSPPSPGCRXPXPTGPGXSAGVRCCCARPGV	6554
QY	129	-----	137	QY	148	-----	147
Db	5475	SSVVRXRCGEFCPGNAGAGLGSXKNTKTLTWIVKRWQIRARRVOXAXKQGGGLVOIA	5534	Db	6555	ANRWSVCSTPRPPSGSRAMTRWSRPRRRPPXSPVTPAKSRRSSAAGPTRGWSVGS	6614
QY	138	-----	137	QY	148	-----	147
Db	5535	DLYIGALGVFPVCVSEVNAVERGLYPABEAHAELGGVALAGDVPPPEMALRAAQAVK	5594	Db	6615	PPTWFTARTWTPCTGTCARPTXVPAHRTTRRRSTRGRPRXPTTARTGRRTCATR	6674
QY	138	-----	137	QY	148	-----	153
Db	5595	RMGSGTPEFDLLYASTWHQPGDPQOAYLORHLVGGDMLALEIROGCGVFSALELAV	5654	Db	6675	CGSPRSPRRPTATGRSSSCPTPWSRTRSTRRWPNAAWRTCSSDRRYGATSPRGPSPA	6734
QY	138	-----	137	QY	154	-----	153
Db	5655	GYLOADNRTSALIIVADNYGTFLIDRWMGPGFIGGDALPPWCCXKPGFARLCSVASKG	5714	Db	6735	PPWAPPTATAXAWTGRSSRTGTSKSCRTPPGSTARCGVSPGCRERPSAATTSTPTRCWA	6794
QY	138	-----	137	QY	154	-----	153
Db	5715	LPEIESLHRGDEPLFPSPITRGRATDFSARIQQFATRSPAFVAMAEIQDHMDVAERAL	5774	Db	6795	RPAASRAATCGCCTARWTTTAAARTAAATPSTAWRSSRPPCWRSRWRPAPRAKSAAPSRT	6854
QY	138	-----	137	QY	154	-----	153
Db	5775	AGAGIGMADVARVSFMYTHAKWSSAEWRPGGCRPGRNSAGRSATCGASHLLSMEH	5834	Db	6855	XXPTRCXRVGRSRSSARAWCGWPPGCRSTPTTRTPPGSSMPRGPRTTSPVWRRG	6914
QY	138	-----	137	QY	154	-----	153
Db	5835	IVRTGELAGDHVLQATAPGLVSSAVLQVLESPOWDEXXIPATTNGICPYPLIARTT	5894	Db	6915	RCWTRASTGSNRPTPAWSPGCHWRWAYPRPVSIGASSDCPPVSVYCTLRCRSTPRGPR	6974
QY	138	-----	137	QY	154	-----	153
Db	5895	PSLRWWSGDAGSPVVTSTPANSCTXRDPATPSARCPTTAGKSTGTSVGRSMRRCGR	5954	Db	6975	CWTPXCRSRRPPSWASRSSAWSCTSTRSPSTARHRRRRRSRSIPASPTCTPWSRTGR	7034
QY	138	-----	137	QY	154	-----	153
Db	5955	FAPAVSWTTTXRVSTPSSASRRARPSXWTRSSAXCWWRPGRHWSTRASRRTPWQGPV	6014	Db	7035	DARWRACAACTRWSSSRPRPTTGRAATRTTRTWCKRVCRRRSCVPGCSTRCASRRR	7094
QY	138	-----	137	QY	154	-----	153
Db	6015	CSPGSAPTTMAPDGWKTCTSTHGRSARRCAPCTASLTRSTCAGPASFTPRARPWW	6074	Db	7095	CDSPRPCTFAARWWSRGSTRXXRWWSAAGWRSASAGTCRPTSSGSCPPSATSSITXDNA	7154
QY	138	-----	137	QY	154	-----	159
Db	6075	RCTPPRACGSASAPWRPGASTCWCCHRARRSPWARPARWRPTAGASRPRRDPATPLR	6134	Db	7155	SRNTRARTAMRPERTDSVTHVSRGRRGERAAPAGPETXGKPEKHASTIVVGRHRMPSGV	7214
QY	138	-----	137	QY	160	-----	159
Db	6135	RPTATAAARAAACWSSSCKRTPYATVTGWCRCARSTRTGAPTASWPPAGRPRSTXG	6194	Db	7215	PLKRPDLTALTARCGIXRRSYWAGIARGCGLRIPRPNWGRWPIELFSPMYWIPSRRP	7274
QY	138	-----	137	QY	160	-----	159
Db	6195	APXRPASLPTLTSSRPTAPVPASATPWSARSPPSTDVTPVRSRARSGRXSPTSGIW	6254	Db	7275	EYRRGLSSGRPRRRLPWETARSAWTSPTSSRTPRAGWPTPCPVTSSATTSSCT	7334
QY	138	-----	137	QY	160	-----	159
Db	6255	RGPPGSGSSRRSSPTRTGSRPACWTATPTSTGTPASTSGWRPGCRGPSVTRTAGRL	6314	Db	7335	WATTTSTASSTSWRTSAATSPRRSCSPGSPTRAGSAWPNLRTAGXPRSWRSRPTPGA	7394
QY	138	-----	144	QY	160	-----	164
Db	6315	CPVSATAAPWPMWSSNRPPRPPAPRRSRPEPCRCRPPRRRSVTARPSRPSGSRALT	6374	Db	7395	TWRSSACTPSARPTCPWPASGRPGTWSRPTPCSGWPTTGTGRSAPRXSTGTGRTSRAS	7454
QY	145	-----	147	QY	165	-----	164
Db	6375	WPRSGTWPITGSPRWSTGRPSWRPAGTNWPPGCARWPLRNPVPSPPGRCRMPAVPSGC	6434	Db	7455	PTSWRXTGTCWSPWRPASTARWTRPANWAEWSWRRGRKSPAPASSVPSSARTWCRTPT	7514
				QY	165	-----	164

Db	7515	TSVPSPSRATAGSRTARSSTRSCCVALSRESFVSRTXSAGRPRXTPPPGRRPTGWC	7574	Db	8595	AAAIERFRDRGFANAGPVLAPDAIARLKAGAERLITRFTDEGLRSDDYWNFPVEGDERPV	8654
QY	165	-----EOGCGN-----	170	QY	192	-----PT-----	K 194
Db	7575	WETTAGRSAAVEACGGGDBGGLREHARRORCGGTGTPGRHWSGLHRLPVRTAG	7634	Db	8655	LYRVHNLEKQDWAPERDILLHRELAQAAAAFVDGPVWPPTAYALVLKEPYRAAEVPHWRDR	8714
QY	171	-----	170	QY	195	VNVDSTIY-----	203
Db	7635	PGRRSAGDAGOADLRRPEPGEGACRRPVHLREARHLRQAAARRGGPRSGQLRGRI	7694	Db	8715	VNVGPRTVCNLSICLDDAGPHNGCLEAVPGSHLLPDDAEVAKVRATGPVVPVPSQGDVV	8774
QY	171	-----	170	QY	204	-----	203
Db	7695	ARRPVDRRGAVHTYQACRPGTHAGVPGRGHQDRAGLHRRGVRQHRDGLLGRGRADRP	7754	Db	8775	VHDVRLVHGSGPNANGSWERTTIVIEYADPAAPPAKXALRRRAPGRHRTAGQGVFRMRIMF	8834
QY	171	-----	170	QY	204	-----	203
Db	7755	QIALRGVQGRRHGRPLRPHFPAWAGERDAVROQLRALSVPGEGRSVAVHHPAAGRPOHPA	7814	Db	8835	TASNWAGHYMCVPLAWALRAAGHEIRVACFPSQERGVAATGLMPVPVLDAPDMMESARL	8894
QY	171	-----	170	QY	204	-----	203
Db	7815	VRGWRQPGLGARVRPRAGHPAGRRTAAGAGLPHRGLGRADQSGTDTATTAGRPGRRLGP	7874	Db	8895	AFVYQAMVTPQSGPRPLPHPTQPMQSLNDFDAGDLRDFWKKSIDAVQRSYDNAVSF	8954
QY	171	-----	170	QY	204	-----FLG-----	206
Db	7875	GRACTREGTRPALLAVRPPRRRLRTPCVVRARPRRHRALVRREPLGVAVERARAVR	7934	Db	8955	GDHWRPDLVVDHIMAVEGALVAALRGVPSVYVSGFIGTETEPGLDLSADPLSCFEKY	9014
QY	171	-----	170	QY	207	-----	206
Db	7935	RHGSGRADSTHXGRARPRASPSLTFFVSRCPDFPGDLMAHCLVTGAGFTGSHVABE	7994	Db	9015	GVWGRDRIOYAVDSPDVAVPPLGDALPLPMRYVPYNGAQAGADPWQLGPIRGRRCVIMW	9074
QY	171	-----	170	QY	207	-----LTA-----	209
Db	7995	LLSRGHRVSLDDLSCGTAERVPEGALFTGSVTDVELVDRLEAEQRFDHVHFHAAFAAE	8054	Db	9075	GNSAGIFGADVPLALRHADAARQGAEEVLTAPQEQVEELGALPAGVRVLRNCPLELIL	9134
QY	171	-----	170	QY	210	-----LLRYAQRNC-----	219
Db	8055	AISHSVKSLNYGTVMVGNLINAALRTGVSFPCFASVAVYGHGTPMRESSIPVPADS	8114	Db	9135	PYCDLLVHHGSANCYMNIGIVAGIPQLSLALNYDTLICGRRIDPAGATLTLGLEATAEKV	9194
QY	171	-----PTTFN-----	180	QY	220	-----	219
Db	8115	YGNAKLTVRELETTMRTQGLPPTAFPMNMYGEWQNMRRDPYRNAVAIFNFQILRGEPI	8174	Db	9195	EEALRVLFDHRYRAAEKLRDGVERRAPAAVAGLLITLVDGGLSAKDAEVDNATE	9254
QY	181	-----	180	QY	220	-----	219
Db	8175	VYGDGQVRAFSYVKDIDVIVRAPETEXAGRAFNVGSSRTNVLLEAQAVRAAGVPS	8234	Db	9255	ARRSAXSGXRRDSGSPSPGPVASSAPTANGSSGVCXPAAWTTCPAGSRTSTRSPTR	9314
QY	181	-----	180	QY	220	-----	219
Db	8235	HPIAHLPARDEVMAVYATATEAREVFGDWADTPLADGLARTAAWAASVGAELRSSFEIE	8294	Db	9315	SGSRCCGPTYXPSRSPSRGRSTTSTWRRPPPWTTIWRCHWRRCGSARPGPTRCGWRWRT	9374
QY	181	-----VPRW-----	185	QY	220	-----	219
Db	8295	IGGERVPEWAQCVADRSLGADRXPAAQVSLGAPDSGGKATVNDLEIRRRVRELEPWN	8354	Db	9375	APASWPPRRRTGIPSTRSPRATCATSTRSARVRCITTRPNGPSPRXPPPTFARWARTP	9434
QY	186	TKLYVG-----	191	QY	220	-----	219
Db	8355	DFVYNGVRYATASTRDYLISQPPQERADAFPAAPGAKRVLELGALEGADTLAMSGQGV	8414	Db	9435	ALPACSTPTDRCAAATTGGWCPSSTRSPDCRSPTAAAPRPGCATSRTCRGDXWPXW	9494
QY	192	-----	191	QY	220	-----	219
Db	8415	EILALGREENLRAEFVMEVHGTNVNELRLADVESMEFAGLGHFDATLCAGLLYHVQRP	8474	Db	9495	TRPSRARSTARPVRXPVSRSSRRSPHSREWSGSPSARRRRTSPAGAAPTSRPPGCRSWA	9554
QY	192	-----	191	QY	220	-----	219
Db	8475	WELLADIGSVDCLYLSTHYWGSSDGMTDLDGTRSTPSAKTTSPRPAASASTCAGWTGR	8534	Db	9555	GSPRCRSPKGSAPCSGGRAPTRGTWPTTAXTVTRATSGVERHDKPDTGCGGADAPERG	9614
QY	192	-----	191	QY	220	-----	219
Db	8535	RXCAPWSRRASPMSCCTSGPRQSATSRRPAARPACPAARTRKEDASMTTATLDR	8594	Db	9615	TQGLGVGRGRGHAGAARGLSAPXRVSARPCRADRPVGRGAVDPGHRSVDRQRRGRHG	9674
QY	192	-----	191	QY	220	-----	219
Db	9675	PRAGAPGRQPAVRGQFVVVFGEDPHLQGAQRHLPPHPERLPGRTSGTSGLLRRRIARQL	9734	Db	9675	PRAGAPGRQPAVRGQFVVVFGEDPHLQGAQRHLPPHPERLPGRTSGTSGLLRRRIARQL	9734

QY 220 ----- 219  
Db 9735 HPPERLRLPLPARDPAGAAARRPVLPQLPGLRRGHQAAPVRVRQELPRGAPCRRQIHV 9794  
QY 220 ----- 219  
Db 9795 LARHRRGPQARGGRADTGTRVARRPGRRLPHIPAQGTGCVRYRTLGRQGPVSSVCLG 9854  
QY 220 -----THSF----- 223  
Db 9855 TWAISGFWSRTEPAVEAVRFDLGVNPFDTAHAYGAGMAEAGLARGLDLLRTRSDI 9914  
QY 224 -----YLV----- 226  
Db 9915 VISTKGGLELGRGVVNRNSDAGFLRANLTDLSRLSGTDYVDVFLVHWPDPVRVPPAETAGA 9974  
QY 227 -----NMSRNL----- 233  
Db 9975 LAGFVEEGLARYYGVSNFTVEEMAEFSVVTPQVAPVFNMLDRGIEKQVLPHCAAAGIG 10034  
QY 234 ----- 233  
Db 10035 VMGWSALAHGVLAGALRPGQVFPDWRAYSPTFQGERFAKLLAAVDRLKEPAERGHV 10094  
QY 234 ----- 233  
Db 10095 AQLALAWLAHPSGVIPIVIGAQLPEHLEDSVRAVDLDEAEURLADELLADAPELDGS 10154  
QY 234 ----- 233  
Db 10155 DOPPAEREVSXGGHDARPPAQLVRPVDRRDHTGLRLPRLPGHRCRRRRDPHAHAP 10214  
QY 234 ----- 233  
Db 10215 GRPAGPALGRAAAVERTAAPRLPRGHGMAGRRHPHWPQQLRPGQLRQRPAPPV 10274  
QY 234 ----- 233  
Db 10275 RPRRRDPGRDGGHRTSGRRRPRGRPGQGGRGRRGPARGGRAAAPAVPRGGLRT 10334  
QY 234 ----- 233  
Db 10335 RRRPRDARLHHORRVPVGAAGRRPRLHAVHPAGLAGGAGRPGRSRRRAAHAVLRRRS 10394  
QY 234 -----FR----- 235  
Db 10395 RAPVRPPPPRRSAFDMTDTTAPSLASLAVPECHIRETGILLPEHVTAPRRQGVLAVRGLIS 10454  
QY 236 ----- 235  
Db 10455 PALESVQEAAGLIDDAMRTRSMHDTIMTLEPHEPDAAPVRIEYVMDKSPVMARLAGHP 10514  
QY 236 ----- 235  
Db 10515 LLLRAMETLVGNFIPTWDSWVEKTTAGAPRLAHRDGMYSDAVATGGGRVIDGIVL 10574  
QY 236 ----- 235  
Db 10575 DHAPEDNCVWAIPQSNYWEDEQVETADRLNATEWDATGAVPAMRPGDALLHNILTLHG 10634  
QY 236 -----VPKYINGTKLK----- 246  
Db 10635 APAVGKORVYVYRPGEVERQLGPHVPEYV-GLKQQVLRACLQRAAASGEHRDEEPF 10693  
QY 247 ----- 246  
Db 10694 EYRPVQYRLWDESPAISGLRPPHEEYRWKTXCAACRCWARSPRRRRARAASPPRRSS 10753  
QY 247 ----- 246  
Db 10754 PSASGSGPIRAPATGPTPICSTATRWPSPPGRRGAGSTSERAPRRTATCSPAPGWRRP 10813

QY 247 -----NTWR----- 250  
Db 10814 RCAAARTXERTITSWTRPGCAACRTARSTVCSPPRSNTXRTCTICARRSGCCGRAAGWC 10873  
QY 251 ----- 250  
Db 10874 CPMASGRSTAGRTCGAGPTDSRPRPGSPWTERSSPADRAVCCSCCATTAGSTAG 10933  
QY 251 ----- 250  
Db 10934 RRAVRSGCCGWPMPGTGCARASITTTSTGSSPAKASRARPSRSTWTSXPPASRASRR 10993  
QY 251 ----- 250  
Db 10994 SRRGTRYDPAPAQSHHVCAAVDPVRLHRRCLRLPLLCGRLDVPQRAVRIRGGAGLOAQR 11053  
QY 251 ----- 250  
Db 11054 RCRLRAQLPLRTGQSRXGORADRGVQOQPOHAADPPAGDAGRARVRVRLLLSPRGPDV 11113  
QY 251 ----- 250  
Db 11114 PAEQHQAVERHGLRRSGRTYRPRRLVRGVPLLPRLPHVQHLQPAVDRVRLPCQLRLGCQ 11173  
QY 251 ----- 250  
Db 11174 EAPVMVVRGDRPVPGRGAVGDAARERPPHAAVLGRRALHAARPLRDAGPADRDQRG 11233  
QY 251 -----KLRKQAPVK 260  
Db 11234 KRRHLQLOPHGAAGEGLRPLPALQERRDRRVLRRRRGLRAHPHAPREVGLRRERTPCQ 11293  
QY 261 ----- 267  
Db 11294 DRGEVAPGRPAABEPVGTAGPAALPRCGPRPGQCBARVPVGRQSPGRGEATGDR 11353  
QY 268 -----KKT- 270  
Db 11354 RAGRTRRVVRVVRPGLEQARRPSAGAEVHERRXPVTRDRRRRSMGSAKAAVAEGLRKY 11413  
QY 271 ----- 270  
Db 11414 GSETALDGPLVVPETVCALLPGNAGKTTAVRILSTLLKPDGGRALVAGLDVTREAAE 11473  
Db 11474 VRRNIGVTQYPAVEEILTGRNLEMMWGRLYHVGRREAQRTEELLEQFDLTEADKRLK 11533  
QY 271 -----OSTT----- 274  
Db 11534 HYSGGMRRRLDLAAGFTVPKVLFLDEPTTGLDPRNRNEVMKVRSMVEGTTVLLTTOY 11593  
QY 275 ----- 274  
Db 11594 LLEADRLADQISVIDAGRGVVEGTETLKSIVGGDRIVVTLAATEDLSAAAQVARTGT 11653  
QY 275 ----- 274  
Db 11654 EPEVDAEARRIDAPVRDRVGLVEIVQLLRGQGHGVVDSLRRPTLDDVFLHATNRAGR 11713  
QY 275 ----- 274  
Db 11714 KEPVGXWPCWEGPAGSPTAGWSAAPTSTGGAI PCXSSTAXCSRXSRWSGGTCSAAPCP 11773  
QY 275 ----- 274  
Db 11774 WRAAVTTASXRCGCDREXCSVSPACWPSPARSPGAXPTGSARCRWRSRRAAASP 11833  
QY 275 ----- 274  
Db 11834 TCSARXSSSSSSAADCSAGAGTRAPARRRRSGCCCGATRSSGSGSTWGCCCCPRPP 11893  
QY 275 ----- 274

Db	11894	GRMCRCPXRCWRTSPSSRRRCRCPAGMAPSRSGTRCPCPPCPAGSCSAIRASRAAPGSPS	11953	Db	12974	ARCALAAGAARRRRGRPDDEAVEGGAGQHAPGAGPGERCLDLRTPTGIRRHVPVRRQPDGRAA	13033
Qy	275	-----	274	Qy	336	-----	335
Db	11954	TRSSWSPGPYCCWCSSCRCECAATGRSADHDGRRRTGDERRTACRPALPDHAPDQGI	12013	Db	13034	GTRPRGLPAGCPQCPQORTPGADRRRAGGRAGGDRVAADRRRRRGPQRRCRHPGRAAAG	13093
Qy	275	-----	274	Qy	336	-----	335
Db	12014	TLPGDVAVRGDRRRHPPHLRPGGGRRGLRGHAHRRRRHQHPSRPWP	12073	Db	13094	HRLHQLAPLHPHAGQPRGLYLQPPGAARPLRGRRRRARDPDPGRPGALPVRRPGGQGL	13153
Qy	275	-----	274	Qy	336	-----	335
Db	12074	RAVHRDRRPFPPRRLLDARRRQPRRLWRQHRHGGRRPHRRRCGLVVAARXPRGGGLL	12133	Db	13154	PALHGAGLAAQGGAGGGRVRAARLPAPGAGRARLGAAGLVLRDHSFVLDRDRADRXGRR	13213
Qy	275	-----	274	Qy	336	-----	335
Db	12134	RRRCAQFGRGAGVQPGPVAAAGVRLRQRLDQPAAGFAGARRPGGGRVRPDGAR	12193	Db	13214	AARPARLRTGAGHVPGGHRAVRHLVRAGCDGGLRGRRRRSHRGHQLRPGDRALGR	13273
Qy	275	-----	274	Qy	336	-----	339
Db	12194	GGRRRRRRRRRRGRRTLPGRRRSHLPLGLPHLPVQRPPLLRGAGGAELPGRRRGRVA	12253	Db	13274	RSRGIRHLRRARRSTDPGAAAGIRRGACGGPVVPGCDRRRTVCGAVPQAAGRSLEVR	13333
Qy	275	-----	274	Qy	340	-----	339
Db	12254	CAXPAHPEHPRGCDGRRHRRDQHADRRGRGVRPRGPRPRRGAAPLCCRHDPTAR	12313	Db	13334	GRGPERVDRPGLTSSVLVQPGARLGVGLFGRRHEDQHIGPARLVGEIGARVMRSAQGV	13393
Qy	275	-----	276	Qy	340	-----	340
Db	12314	SDDLMAALAVITALNQALHDEMARDRCVFGEDVRIGLTQVAKGLHERFGDGRVVDTP	12373	Db	13394	HHGLTRGEDGHPGGRVHRVGHQPPDVTVVVDVVRQHGTEVRAGHIVQAAAHHVGLVOR	13453
Qy	277	-----	296	Qy	341	NPFCBPS	353
Db	12374	SEQAFTSLATGAAMAGORPVVEYQIPSLLYLVEQIANQAHKFSMTGGQVEVPVTLVP	12433	Db	13454	GEYAPAPAPARVHVALVLPGRCAVVRLQNGVPIEVELVRDVQQLGRDPAPLEMR	13513
Qy	297	-----	296	Qy	354	SEFMK	368
Db	12434	GSGRSGMAGQSDHPYSLFAHVGIKTVLPAATASDAYGLLLSAVRDPDPVAVFAPSALMG	12493	Db	13514	TLDLQDGEFLRAHDLRXPGRPVGILAAAGDEVVARSEVGLGEQAIQVDTAQLQRLRVHE	13573
Qy	297	-----	296	Qy	369	-----	371
Db	12494	TVEVSGELGPVPLGSARIHRTGEDVTVATGCVHVALAVAEAMADEASIEVVDPRTIY	12553	Db	13574	PRQCESVLEGCALLVDHSHFGHPTAAPFVRHLRXTIRSSPEYIGIFSRIVKNLDCS	13633
Qy	297	-----	296	Qy	372	-----	387
Db	12554	FVDWETIRASAEKTGRLWJDDANRMCVGFGEVLATAAEQFDLTARPRVTRPDGAVIPY	12613	Db	13634	EXRXIHGRRALPLMRVXSRRDVCKIRERVAMTGHSAIALDVGVGVYVDEPFELAWQ	13693
Qy	297	-----	296	Qy	388	-----	390
Db	12614	ALVLDQALLPDAQLTDAIYAVLXKRGTPXLERERRXIPSVGRSGXSRSPSGFTSCAGW	12673	Db	13694	TFDLAVTADPSVTLRDFIEDTERFYLDGKTPAERPGIILSAAADESWVRVRRAWRELAQE	13753
Qy	297	-----	329	Qy	391	-----	390
Db	12674	TGVSAPCWATPASTWRSATSTTRKTAVSSMCARFGRPRXRSSATTAIWASPGRTTSAS	12733	Db	13754	IPGSTAAVTALAREMPTTVANQPPECADVLAEWGLTAVLETVVLDCFEGVAKPDPRLIG	13813
Qy	330	-----	329	Qy	391	-----	390
Db	12734	AAPWSPARTCSSSTTPGRRPTWCADSPNWRNGRSTASSARCSTATTTPRSTSTSS	12793	Db	13814	IATALRGRPPGELLVWGNRVDHVDAPARSLGCPAVFVLPEDGYRRPPGVHPDLTRYRAL	13873
Qy	330	-----	329	Qy	391	-----	390
Db	12794	GRTRPCGMASSTPSPATGSPIPRRRCRRGGHGRWSTPCWVRSSPGPRCCARGCST	12853	Db	13874	RAVRTVPDGDVPRVASLADLAASPLARLGAIXPRCTCCPGTVRXSGRRVSSDPGXRP	13933
Qy	330	-----	335	Qy	391	-----	390
Db	12854	RCCTPTTRPTCAAGPAGGAGGWPCTWTWASSTGAAGRPWPARTAGSTCGATATTSSP	12913	Db	13934	SPLTARARRPPSTGRSRRWSAAGPPDRWRRPTSSTSWPPASVRSSPNVGTWSRENDCW	13993
Qy	336	-----	335	Qy	391	-----	390
Db	12914	TSTGTRRRPPGSGAGSWTSAWAAASAAASRRPRPERGRPSRRRCAGSPTCPCGPAAGTT	12973	Db	13994	SRCWRRCERDTARCSYWPAPAGRSTRPSRNRTRGRPARPGTFTGTPTSSASNASCACT	14053
Qy	336	-----	335	Qy	391	-----	390
				Db	14054	RMSPVPMXGXMSTVPASAFCAAGSCVCCRTGXRRRPPGASPCGCSAIRMWSGMTSTSTSP	14113

QY	391	-----	390	QY	399	-----KTTT-----	402
Db	14114	GSCSRCAAGSAAAGCPPWXTSARACPLRXAACWRSQVXPAAPSRGAGSGDGPSTGRATG	14173	Db	15194	PGCSPIPRSRPPGSGTWTITASAPARPSTTSAGTSSRTWRTPTCRGSTTRRXARAPVGS	15253
QY	391	-----	390	QY	403	-----TSPS-----	406
Db	14174	WMSGHPAPNSTGGPPSPXPRVSGSAGSVCTTTPAGPLPPRRARRSDALDQOTRQTHPRTE	14233	Db	15254	PAWSRRTRTSPASAPWWTSGAGTGRCCARTPRCAACCTRRRPGSRPPGRCWRRRAS	15313
QY	391	-----	390	QY	407	-----MGFORT-----	412
Db	14234	RCEAPSPYPSGEARHORCDQGLDRPIPSVGLVTRXRPVAGSRXSWASILPGASLTGAAG	14293	Db	15314	RTARWRPATSSPPYPRAVTSTCSRASSMAGRTSGRPGFWRTAAGHCRRTAGSSWSSTCCP	15373
QY	391	-----	390	QY	413	-----	412
Db	14294	IPAAAIKRAFSAASNRTELSSGRASAWACAWARSSCVPTTASWMSASPATSSAKATA	14353	Db	15374	TRCPRTPCRPTTSTTCWSTATGWSARVATSSSARREARSRKSRRWPEPTCGXSRPC	15433
QY	391	-----	390	QY	413	-----	412
Db	14354	PSTCSYSAVARGSVIASASSKKAALPACCWYTPDTRVNSRWXALRXSDSNAPASSVTS	14413	Db	15434	PURLTRSAERPPREFHRAREPHCGIGPNGRPTQOQTLVNSLYCLREQXFTWALPVFPFVT	15493
QY	391	-----	390	QY	413	-----	412
Db	14414	LFTVASTDVAASMPXLSARWTPVDSGSRKQASPARXXPGPWKLSATYKSCSTRTPGW	14473	Db	15494	GDTHYSRLGDDSVXASAHYGNHWPWHLPRASTHRIRIXSLRNGARAGPAVCPRRSAPAP	15553
QY	391	-----NKTASDSN-----	398	QY	413	-----	412
Db	14474	XTPFAMRSRTVGDSTSEASIASRSACRRSRNRSMWQTKPTLYRSRSGITQYQPSGA	14533	Db	15554	GHPSGERRLAGDEPDRCRAVLPHRDLOPRRPVVRPGILLEAAHRCRGGDROGCAADAGG	15613
QY	399	-----	398	QY	413	-----	412
Db	14534	LMTLMPVVGELSGSEPPCOXQIATLRKISCGRRFPFCLEMTPARPEASMTXSNSSLTSS	14593	Db	15614	HPHGGAAQFPDHOQRRPALLSGLHLGQVGRKAGVPAERGHALPPAQPLRGGQVGRALHG	15673
QY	399	-----	398	QY	413	-----	412
Db	14594	PVSGSHSAPAAARSPNSNTRLTVPVTSAPCLTAPFRSTWSKSRSTCQERTSGPKIEG	14653	Db	15674	AELPRVVRDVRQDPVQPRVTDPCGVHQAQVARGRREAGPDQAPAGPQGGGTRLG	15733
QY	399	-----	398	QY	413	-----	412
Db	14654	QSIQSSLQSRSTPPLTVKPRESTSSAPTRSMVVRHEAIDS PMWXRGKVSRSNSTD	14713	Db	15734	LRRGLRARHDDARPCRAGPRPRHRDHPPGARSGGFLRARRAGLARPRGGGLPAAAPG	15793
QY	399	-----	398	QY	413	-----	412
Db	14714	MPALLSKEAANDPAGPPPTTSTFLDMELPTDNGSTSRRIADRNEHPARVIVKYFSECSR	14773	Db	15794	GGTALRRPEPRSEARLEAREGLVRGTGRDGRQRPAPAHROXGRRRRRGRSRTGRSVSL	15853
QY	399	-----	398	QY	413	-----FTD-----	415
Db	14774	DALVESVLLHARLCSIRFGISAELEDRPVKERDGNPRTDRGALESFVSDGFIHLPGAPP	14833	Db	15854	PAGQVGLARABALPSEAVLSDVAGAEFIDWFARCSAAAQCRVTRAPLTELRRWRFDAT	15913
QY	399	-----	398	QY	416	-----PLM-----	418
Db	14834	AELABEACALLWRQLDMDPDDPATWTREVVLGVRDDVFLRAANTPTTLHAAYDQLAGRG	14893	Db	15914	GNLAHESGAFFAVEGLRVRTSYGPVPEWSQPIINOPEIGILGMLVKEVDGVPHCLVQAKI	15973
QY	399	-----	398	QY	419	-----DYLDS-----	423
Db	14894	RWQPADAGRHFVGAVSDEAARGDLRLAHRRLQPLGRGRRRPGLVRXAGRDPAGLRQD	14953	Db	15974	EPGNHRIQLSPTVOATRSNYTRIHOGRSTRVLEPFTKAGAGRTLVDVLOSGQAWFLRK	16033
QY	399	-----	398	QY	424	-----LLFLDBI-----	430
Db	14954	LPLQVVPRPGAAAPAPVLGHRRRGRAHADPRRLPGRAAAAGAVWRGHLPGRGCGGTG	15013	Db	16034	RNRNMVQVAEDVPAAKGYHMLPLTELRLRLRIDGLVNMMDTRTVLACLADTTPASRPVP	16093
QY	399	-----	398	QY	431	-----	430
Db	15014	PAAQVCHGEGRGLSVPLPGAAGAGQHRHPALHGPAPBAEPAAGAARTPARRAVHPGRA	15073	Db	16094	VGDVGAALVLSSTGGQTSINDMGAVLRWFTEAKSRHQLTAHRTPLRDLPGWRGTPEQITH	16153
QY	399	-----	398	QY	431	-----RNFS-----	434
Db	15074	GSAPFGQGHATRGVTRWRSRPWRPPGRTXSGWCSAEWPPRSWAMRCGCGCPTSRGR	15133	Db	16154	EDGRHFSIIGVTAQVDREVAEWDQPLIFPQGRGVVAFVIVKIEGVAHLLVHARPOFGLL	16213
QY	399	-----	398	QY	435	-----	434
Db	15134	ANGPPTGWPTSRANPLPYTGCAAWPRSECCASPSPACSPXHRSGNCVAPTARRPSTHW	15193	Db	16214	DGMEMGPTVQCVPENYPOGPRPFLDYVLTASPERLLYSALAEAGGRFYHQSNSYLLVEA	16273
				QY	435	-----LRSPTYVNL-----	443

Db	16274	DDFFVAVPDDYCWVTAHOLTRLLRHGYVNVARSLLACLQSWXPHERASRAAGHSRL	16333	Db	17354	QVFARRPIVRVAGHPAHQFPMRPPPGWQDAGTRXEGRGDGAEXAGGGPGRAASRGAG	17413
QY	444	-----	443	QY	448	-----	447
Db	16334	LRHRPTYPAGRAEGAGDDBRRLPERGQPCORRAVRRRRRLRGAAGEPCRRGLR	16393	Db	17414	GGRRSTRRRRTGGPTDPXATGPARAGHPGPGVAGPHRGOAAPGGGGEGRVDPGRPG	17473
QY	444	-----	443	QY	448	-----	447
Db	16394	LPAHRTAPRLDRARPARGKARAGEAADRLPGHRDGGPPRGLPRPDADGEPDLHPALAA	16453	Db	17474	ARRRLVPAAPAPGRAADRRPDGRRPADRRCTGQHRRTHRRGQARAPDHRITGRRAARA	17533
QY	444	-----	443	QY	448	-----	447
Db	16454	RRRARDAGRHRXTALTALHQVRVPATASQRPVPAGAGRRQPAQGRVRSARRGEHVPRA	16513	Db	17534	GGGPRGTGPRRRGGGADBEARPAACRGAGGAPAAHAVRGAGRTGTGAGRGGDPRTR	17593
QY	444	-----	443	QY	448	-----	447
Db	16514	RPRCGDAADNGBRSRRGRRAAVCARRPHGTRVLRLRLPLRVHAVGERGPDHRRP	16573	Db	17594	AGRPRRCATAARPGLRPAAPGLRGRRPGRRLRAALRVLRPARRALGPRPRPGRGVGP	17653
QY	444	-----	443	QY	448	-----	447
Db	16574	CLHPAARLPHGSAGTRRRGARARPAARRPVRHRPVLPRGRHRRVVGGAQRCHTGTPI	16633	Db	17654	AGRGAGQDPCARRHAPPGLTARLPLGRHLRPRPAPDDPRRDGARGQDDPGHRTTRAP	17713
QY	444	-----	443	QY	448	-----	447
Db	16634	GKSGAGGTAAORQLVRRSRWPMASAFSLVGRGSWAVSCARTXPRPBRWRRSPGYR	16693	Db	17714	SRGAEPLHGILSGORADQLDPCDREAOCAARDTAARPRPAGRVRRLEGARRRGHLLRR	17773
QY	444	-----	443	QY	448	-----	447
Db	16694	TSPCRAGSSWTSPPRRPGNWTXTSPSARTWSTPSAATGGSPNGTWRPTARCRPAACW	16753	Db	17774	AAGLPGARRRGHARRRGRLREESDPPLHGLRLRVGALRPRAVPDPAHGEPRRGVGP	17833
QY	444	-----	443	QY	448	-----HRRR-----	451
Db	16754	TRCGGPRAVPTSTSAQCWSTARPRPVNRPGPERRPGRRPPTARPNRSRRARPQWRPPRA	16813	Db	17834	QFGADPPARAGGHRERRRGGLQGVPGGRTGLSAPQAGRPHTTPGAPAHGRVCGAQH	17893
QY	444	-----	443	QY	452	-----	451
Db	16814	SWRRACGSFTWGRAPRBSACSAGWPGGSPRRSPGTCSSRSWSCAPTATTSTCATS	16873	Db	17894	GGRGGLPRGRARRQLHGDAGGVGRASGAVREGAAPARDXKGGRRERAEGRVLRVPGRTG	17953
QY	444	-----	443	QY	452	-----	451
Db	16874	RTRCWPPAPVRGWSPSDAARPCRCGSSICSSRXAACPPRSGSFPPPPRABPOTGGS	16933	Db	17954	SSTRGAHGRARGGSRPAGRAGVRRGASHSCAGPGGGKSGHAGPGGDRRTTGHTG	18013
QY	444	-----	443	QY	452	-----	451
Db	16934	RSTPGPANSSAGPPSARCTSRXPGCPTRCAGRQSPSTRSGRPRPCQIPVCRAARPP	16993	Db	18014	ADGADGAAQAGTAGTACACVRADLGQAAARPPTAARRLLRRARSRRRTARADRRSGGPQH	18073
QY	444	-----TPPE-----	447	QY	452	-----	451
Db	16994	GGREFDDRAQAARKAKRPLDRTPPHVFPVLKVSPPGTPPTPAPADACRPPGATP	17053	Db	18074	AGRLRGDPLRAAAQADRRTPGGRPCPHGPGPLGPRAPPALHGRGGRTPGADAGEQVPPV	18133
QY	448	-----	447	QY	452	-----	451
Db	17054	GRTGSARVPDGSPPRGSXAGRCPSPRGRCTALRPGASRRTEADAEPCDQGLLGVRN	17113	Db	18134	RLLRPQGRRRVDTGRRRHLGPRPGPHRGGTGTGVGVSSGRRAYSVVAEVKARQAVGL	18193
QY	448	-----	447	QY	452	-----	451
Db	17114	EVRPGVVRGARGRRRLRHQADRRAGFLPHLAVRDTGTRGQGORPVPEGEPPRLRL	17173	Db	18194	TWRVGPSSGVFLESRRMMVMWQORATSTHWPPSLPLXLDLNLHSSAVLIYLPDPLAGSSFP	18253
QY	448	-----	447	QY	452	-----	451
Db	17174	RRRLPARRPGEGGADPGRAGGRDPAAAAVGRRRRHGCRARSPRSAQRHHDVGEADQG	17233	Db	18254	AESWPVTAASHTCAPGWTTPRPGQKPTGCHVGSRCRPGADRASPGGXRGPRVRRRCW	18313
QY	448	-----	447	QY	452	-----	451
Db	17234	DGLRLRPPAAQKRLGGAESAPYPRPHGGAERLGSKARHLRHAGTGGRDQGTDRQAPLP	17293	Db	18314	VETGGRAPVRRTRVGGAGRGGLSAPVRCTAPAPTXXSSSSAGTCGAGSPGVRCDPPS	18373
QY	448	-----	447	QY	452	-----	451
Db	17294	GPRSRPALPGVVGHDAARLRVSGTAGAQDRTGFLSRPISAFNKSAPRAPREPTXSG	17353	Db	18374	WLPARPRRTAVRABAGPRGRSGSAGRAGRRCTSTGRWPPAPACTAPTAGTGRSPA	18433
QY	448	-----	447	QY	452	-----	451
				Db	18434	CSRPRQPTCPCRCGRVPRPAGRVRAAGAAASPVRSRTWSRRRGPLCPGNSAXSPPCGRPA	18493



```
QY 452 ----- 451
Db 18494 SACPPSAWSTGSSAAASWSSCWPNPAPHAPIRTCRCPVQNPSTGARTDKSQRCATFVEX 18553
QY 452 ----- 451
Db 18554 TTQTNSTSATSTHATADPSPLPATYPPGHSSTGASPTGTGTTYSSAXASWRPTPSST 18613
QY 452 ----- 451
Db 18614 ASRPAMVSPASANTATSSASRTTAERAYPEPTTTKAAAVCSARWPTSGAWASER 18673
QY 452 ----- 451
Db 18674 PAKSSGSPSCRALVQLWRPQSGXGTGEAQXVDMPLPTPPRRPHKRTTDXGGPLRWQX 18733
QY 452 ----- 451
Db 18734 CVIPGEDWPPSAXRPPRPSRASSPLSPRPPRAVSVSTRARQTPWMTATXSTRRTTPV 18793
QY 452 ----- 451
Db 18794 RVPRPDSSRSTAPTSSARTRRDXTRSPRRRPRRPPRRSPRSPWRSRTACSASTRP 18853
QY 452 ----- 451
Db 18854 RPTPPCAWTAWTRTSPXTARTPTTRPGREXRTSSTPASASPTATSAAGPTATPS 18913
QY 452 ----- 451
Db 18914 TTTTTPRTATARTWAPRSAPTATSPRRPRXSASACKTTPARAPPPRSPASTSGPT 18973
QY 452 ----- 451
Db 18974 PSSRPSPTCSAAAPTRPSTRPVATWPPASPFWRPATSRPTPPRGHPHASRPSRSAR 19033
QY 452 ----- 451
Db 19034 RPARTRPATPTTAPSTSSPPVPSRPGTQATRRTPSPVRRWRPRTWRAPRSTSP 19093
QY 452 ----- 451
Db 19094 TPRPRPRPRXRP PPPPASPTPARARPTGSGTTLARASRTPVTTRSATTP 19153
QY 452 ----- 451
Db 19154 RSSPRXRSPASPTRPRSPXRSTSTRTSATSRSSKSPPTARRTRSSRTAPAAVTTST 19213
QY 452 ----- 451
Db 19214 PRTEXTPPRRPTARGNGXATRTTTPGAXTPGRCSSPPXPVLVLAARHVRPTPTA 19273
QY 452 ----- 451
Db 19274 RTTHVRPHAASGKVEARWAGWVGASPOLRAGGVPSYACCADGARTPTGQSRTAGR 19333
QY 452 -----VN----- 453
Db 19334 GRVGGCPAVTGVNTRHLAOGCRSDGHPARPRTTQPKATRTTTPPQAAAGWRT 19393
QY 454 ----- 453
Db 19394 TPTTAGSPSGSPRPASPPAYHSPSTRTPVPAARPCPNQEHGTRVHTSAPRTTAARVP 19453
QY 454 ----- 453
Db 19454 GAAPRAERAGVPATVAIPARRTRGCTGVVPPLLPQLRPRTQRRRRRRPGRHP 19513
QY 454 ----- 453
Db 19514 LLSAGRAGAGTDRAGADQAGRARQGHGDSFGALLRTRTPVDGVAVHGLHRRRRRLQ 19573
```

---

```
QY 454 ----- 453
Db 19574 GRRRRGHPLRPQGRLLRLGDRHPGRLRGGRGALLRARRRGVGRHPPRRRLPGRR 19633
QY 454 ----- 453
Db 19634 RTHPDRGOVLBALRRYAMDTHAVPDADVYFIPTRQSWRFGDRREITCLFGATQDDGKL 19693
QY 454 -----LSTNS----- 459
Db 19694 TGSLSRSDPTTLDADQVAFLSATNAVDTALYEBPEEYAEEDLAANKAWAKDVHVLGEQIE 19753
QY 460 ----- 459
Db 19754 GCADTAGRAPPSPSPINSRRWRTPGTRGRRRRRPRTRTRTTTADTSTSTAPRSPR 19813
QY 460 --LWVW 463
Db 19814 ARLWAW 19819
```

RESULT 14  
ABP76680  
ID ABP76680 standard; protein; 19938 AA.  
XX  
AC ABP76680;  
XX  
DT 26-FEB-2003 (first entry)  
XX  
DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 4.  
XX  
KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
OS Streptomyces viridochromogenes.  
XX  
PN WO200268436-A1.  
PD 06-SEP-2002.  
XX  
PF 24-AUG-2001; 2001WO-EP009815.  
XX  
PR 25-FEB-2001; 2001DE-01009166.  
XX  
PA (COMB-) COMBINATURE BIOPHARM AG.  
PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
XX  
DR WPI; 2003-018650/01.  
DR N-PSDB; ABZ37516.  
XX  
PT New avilamycin derivatives, useful for treatment of infections, and  
PT nucleic acid encoding avilamycin synthesis enzymes.  
XX  
PS Example 1; Page 68-301; 319pp; German.  
XX  
CC The invention relates to avilamycin derivatives (I) with antibacterial,  
CC virucide, protozoacide and fungicide activity. (I) are useful for  
CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
CC or veterinary medicine, particularly where caused by Staphylococcus  
CC aureus. (I) are more hydrophilic than known avilamycins. The present  
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-  
CC ABZ37516)  
XX  
SQ Sequence 19938 AA;  
Query Match 21.7%; Score 531; DB 6; Length 19938;  
Best Local Similarity 1.1%; Pred. No. 0.016;  
Matches 221; Conservative 83; Mismatches 126; Indels 18858; Gaps 70;

---

```
QY 2 GRKEMVR----- 9  
|||:|
```

Db 422 GRKQMPRTAPMATPLTGAGRPXGLWEPXGWMVAYGFGAPGAPCGYGDGGYWPACFCP 481  
QY 10 ----- 9  
Db 482 WGCSDPCGCGPGGGMDGTGTVLLVRAAGPGGNGWTTGRMVSGRGGTGKGPRWXSR 541  
QY 10 ----- 9  
Db 542 GGGARAHACGLWVGGAARSIGLRCGSPRRGVSVLGPAAPLQGEVPRVDAGHCGRT 601  
QY 10 -----DVPKMFVLSISFLLVSFINCKWMSXAL----- 37  
Db 602 PDTSPSRRTLRAGGTRPIGAACVAGDTP-----CSKGLGGGP 641  
QY 38 ----- 37  
Db 642 YPPRAGLDSLAGGVPDVGPRGWPGRDVPVPCREDDXDMGSRARTAAPRLCARCRS 701  
QY 38 ----- 37  
Db 702 SRCRSPAVSTCRWPPRRRSPSTWCXCCPNCRRTTXXACTPCRRGRSAGPGGRRCTGR 761  
QY 38 ----- 37  
Db 762 CGPLRGPRAACRRRRRPSGTRPWCXCSHRCRSAGPEGGAAGRRGTAGWASPCR 821  
QY 38 ----- 37  
Db 822 GWXRRWRRTSAPWRPGTGWPRGRRGARRPPRAGSPSTYRRWCWSPSLSSRPXW 881  
QY 38 ----- 37  
Db 882 TNRGRGRGRSWSRPMPSRRPXPWPRXRVTVSWRRWSTRCRPPRREXRRRPMR 941  
QY 38 ----- 37  
Db 942 CVRPGCRRRRRGTCWRRPAXRRSGRPSRCRRRPGWCPGRSCSARGRLRNPWNR 1001  
QY 38 ----- 37  
Db 1002 RRCRPPCRACRGRPGRCRCRWRRSRRRPGRRSRCXGRCRCXRTFSLPARPSG 1061  
QY 38 ----- 37  
Db 1062 RCTSCGGSSGPGPGPRAGWGSRCXTCGSPRRRPPGRRRAPPPRRRPSRSRS 1121  
QY 38 -----YNRPMRG-- 44  
Db 1122 GPSTCARCRRRTSRRGRVPRPNNGRGPXPGRSCRCRPRRLPRPTGYGPRGPRMGR 1181  
QY 45 -----I-VLS----- 48  
Db 1182 RRGSRGPRWPXRGPRVPFAGYDASLPSESSSIGGALVCASTRSGQHVSILCSASA 1241  
QY 49 ----- 48  
Db 1242 ERLRPOELHKGPARTRPRLCRAFARPRPTCRPARPGADRGLRRGRRGVPRFC 1301  
QY 49 ----- 48  
Db 1302 RGPRGRRRHSGREENHGRAGRGGRWSPTRSRRGVRRRPGPSRRGSPXVTRRSSC 1361  
QY 49 ----- 48  
Db 1362 GQRGWIGCVRTGRAGGVRLRRSRDECEAXLAFISACTRTRILYGSTPGYGVWGV 1421  
QY 49 ----- 48  
Db 1422 RPAXTPRSCRGRTRCPGWWASRCRPPARRRARRVPMAKRSSPTRPSRRNRGGGPC 1481  
QY 49 ----- 48  
Db 1482 CSNSPRRWTNSSTRARWLPGRSRTCRRTSRRCSRSSTWCTSGRSPFCAGTTTPSPCAPCS 1541

QY 49 ----- 48  
Db 1542 RPATTPARSASAYRTRTRTSWKPRRRVASNARRSCTASARRXXASSCGSCTNSGGR 1601  
QY 49 ----- 48  
Db 1602 KSTATSSATYASAPNSRGSPSPFYPCAPPHTPDWTAPSWRRPITSTWPTPRANVAASG 1661  
QY 49 -----KIGVKYLDQLKLEILRQ----- 65  
Db 1662 FLARTCPSPWRASMRCCGHRPXLRRRRRACKXTVWEINEHRTXVVQVELQORGRPVRRG 1721  
QY 66 ----- 65  
Db 1722 CPLPPHHHPRLQEHGPRRPHPGQPHRLAGLHLRHHRIGAAATGTHAPNRACSSSTMWRA 1781  
QY 66 ----- 65  
Db 1782 RARSEMSTASGVDSATALRRAXASNRDLFSSIRTRRSSTPSVQSRRCRTTERTWPMRWA 1841  
QY 66 ----- 65  
Db 1842 RATAWCSTVGLSCGSORITTEAACMLRPTPPASICASSTAXPGAAAKSSTSCCRAGTIL 1901  
QY 66 ----- 65  
Db 1902 PEIGPNTRPARGPGTLRGPVRSVGAGVSRTAATISSWPGVGVVAALSTAGPRASAVARPSS 1961  
QY 66 -----LETTISTKYN-----VSQK----- 79  
Db 1962 DFGSSCGPAPGCPAASVGSARTTASTSRNAENTTTTFSPSAASSAISRCTFADRSGR 2021  
QY 80 ----- 79  
Db 2022 AYALRIAMKLPARTASRYASSAVLGGSHSVCSWAGSAGSVRTSCWLRXXYGTATALRNA 2081  
QY 80 -----PVKNLTWN----- 87  
Db 2082 LEPATPSCSSMASASSGWIRTRLRNSSTRFSMGVPVRNSTRSQCSAHSETAMERRVRF 2141  
QY 88 -----TEPP----- 91  
Db 2142 FTXWASSTTSMPTPPPGAGKPRSASKVVTATPPRPFQSANASCRSGPCGVTRSLRFS 2201  
QY 92 -----OYVILAGPIQNYSTYLV----- 109  
Db 2202 VSRVQLISTLAGQTTMKWLCPSAARCASAMAWIVLPSFISSPRIVRWRARAKRVPKAWX 2261  
QY 110 ----- 109  
Db 2262 PRSETRRCVSSAWVLTRSATSSGRNPASAGSGARPESASGRPXYSABRSSKSTRAATS 2321  
QY 110 ----- 109  
Db 2322 BFCRRSRSTEAWASGGTSPWSACSVSARISATASARPARAARPANSMRSRPAAGPASAS 2381  
QY 110 ----- 109  
Db 2382 SGPSFVSAPPASTARRASGPASTSTCSATTSAVVRGSLSTSMCTAMLSTRASAIRWAA 2441  
QY 110 ----- 109  
Db 2442 ATVMSAVCPTPSRWRSRVPEASSTWPTVRNPPFTSPRRSLPSVRSDSRATGMTSSRR 2501  
QY 110 ----- 109  
Db 2502 ASCSWIGSRAARTASTARAXSSTTACAASXSRPPRPARATASRLRPIAPPLSPRSRLPPG 2561  
QY 110 -----FD 111  
Db 2562 RGSRRLLVRGMPGDDPRPSEHLPALRPLAGGPRRPEGRDRPAQRSCSILACSRSS 2621



Db 4782 RATAGSQVRSRRRPARRSVPRPSRRRAAAAAAPCPRGWPFRRTPAGSGPSRGAPR 4841  
Qy 145 ----- 144  
Db 4842 PASPPSRHRPSRSGMSLPRPRRTACSTGYVRPSGRRSSSSCPTGSRFGWAMKRGVRPVLIAT 4901  
Qy 145 -----VPSM----- 148  
Db 4902 GVCTKGMHRKASPAFPVADLSGRSRPTSPASRXVPSAPYSGSSSGGTSRWEPARISVGAS 4961  
Qy 149 ----- 148  
Db 4962 SSPVSENRSRSSARPRDHTLXRKILXSXGGITSSSPDQSRASASSARKLASCOPXSS 5021  
Qy 149 -----TCLS----- 152  
Db 5022 VSSGRFVTGNRTGNVPTCVSGLPPALPGELVVGVRGVRGGAGQEHVVVAHPQDPHPPRP 5081  
Qy 153 ----- 152  
Db 5082 GGRVVGHVOLPPOQGARLPQELRGERTQVDEAVGDEALKCASVSTRVPIITLLYRTVAE 5141  
Qy 153 ----- 152  
Db 5142 LGTNPEADATYSRMMKDXRFHXSVATTLGEVFNHYPCRMFISIRNAPRRETXIISGELHV 5201  
Qy 153 ----- 152  
Db 5202 QKSARDRWAGGIHGLLAQGRHVREAI GTGDL PALPHRRVDRLLVPHHRRPRRRAGG 5261  
Qy 153 ----- 152  
Db 5262 RLPRLHQERRAAALGQGLDRLAEDLRSRRLLAGRPRLRAPRAEERGQAGRRHRG 5321  
Qy 153 ----- 152  
Db 5322 RHQOARVRRPRGWSGVRPHRXAQGRVLRHRRRLGPRRHLQAPEPPPPARDLPQR 5381  
Qy 153 ----- 152  
Db 5382 RDLGLLAGLAAQFAHRGHQRHORPCRLVLGDPAAARGPVQRRFRLPPGVPVGPAGPTRR 5441  
Qy 153 ----- 152  
Db 5442 PGDAGLAGRGI PDRAPHGERLPPAGRTGGAGILLRRRLQRPRLPRRRRRLLPAPAA 5501  
Qy 153 ----- 152  
Db 5502 VHRPPGALQRHAGRDVRAHGGQGRHGGGGRGVLRVALPQRLPAAVHPGVRVRPATGRQ 5561  
Qy 153 -EMLNVSKRND----- 162  
Db 5562 GRLLRAGRDDPGARHGRVAGRRRGRLRTRGTGRHPRRHRHARGPRPAGARPGRRP 5621  
Qy 163 -----TGEQG-----CGN-- 170  
Db 5622 AGQLRPPVARGGRERFPDGGGRDARRTGQRSGRQDGRPRPLRPHRGSLSFGDPQDAGNRP 5681  
Qy 171 -----FTT----- 173  
Db 5682 VQALITALMPGFTTRVWGWSPASLRPRMLTRLTVXRVPSSASRRKGPGRRRRGAHAPS 5741  
Qy 174 ----- 173  
Db 5742 TAGHPRSGGWRPASRVRRGVRHPASCAGRRRTVPAPSRPRSRRSHRTRSAGRSASKWA 5801  
Qy 174 ----- 173  
Db 5802 RPSRRSPRAAGRRRPGRAARAGTXRRRRGHSPGPHADRTPARARPGRPPFSASAAHR 5861  
Qy 174 ----- 173  
Db 5862 TRAGRAGDRBHCSTARPBRGHPSAGTARVGGAWRARRRRRSGSGWSEGTAVRPRARR 5921

Qy 174 ----- 173  
Db 5922 PPRPSWVPRTPGGVPLWTAPPAPRPVAVGRPRGRADVRRRGDXGRRRPGSRRRCRPPVPVG 5981  
Qy 174 ----- 173  
Db 5982 RPSRRPPAGTPGXRRRPARAGRQEXAMPXSAGIRXNPSPRPPYCPRAAGPPWPSORTEARQG 6041  
Qy 174 ----- 173  
Db 6042 RRGQVGGGDPRHVPPRRHGTDRAGAVGPGEVRMDARWPAVLAHQYEHRRAAOQTGGC 6101  
Qy 174 ----- 173  
Db 6102 HVVVDVAHHQQLAGGPAEPGQDAAQOPGIGLGDPPFEAVEDHRLQDGGQPPFFREYVGALG 6161  
Qy 174 ----- 173  
Db 6162 RLIRDHRRHFAGEGRDRENRTNLLGQPPGSPNDPALVGRCRRENARSLRRFPVQI 6221  
Qy 174 -----FNM----- 177  
Db 6222 ETLGVLDEIAEGNTRISGDRKVEGPLKPSQFERLVVDDASDIQNRNRTVSSHNSNLPFYFT 6281  
Qy 178 ----- 177  
Db 6282 YVAPRVLHAHQWERRPTAMDSPLPARVXSIHYSPAKADAI FRGGTYRSQPMPDERRRMG 6341  
Qy 178 -----FFNV----- 181  
Db 6342 AVETVIDEBHRAAFQADGFASLPRLVDDTELEWLRGVYDRLFSEHADFTTGDFDIAGRQ 6401  
Qy 182 ----- 181  
Db 6402 DSDRPARLPQIVRPEKVPPELVESAHFQRCRAIAAQLLDIPEDELDYGHAILKPPNYGA 6461  
Qy 182 -----PRWNTK----- 187  
Db 6462 ATPWHQDEGYMDPRWRRRGLSIWTPLDENATVDSGLHYVPGAHGLPVLPHRHIDDDDR 6521  
Qy 188 -----LYVGTKVNDVST-----IYFLGLT----- 208  
Db 6522 GLVTDVDAVPAAGVPVPLASGEAVVHALRAPHYAGP---NLTDQTRRAYVLVFWGAAEEVA 6578  
Qy 209 -----ALLRY----- 214  
Db 6579 DPEPRWLDKDXRABARAVHPLRASAPHQVAPASSURKCYSTRSAIVTWPHNRTAARR 6638  
Qy 215 ----- 214  
Db 6639 TSSDAGSPARVRRASSGAKSVYPPTTSPTEGTISARQPLPVPTTGVPLTAASSETTVAPGS 6698  
Qy 215 -----AQRNC----- 219  
Db 6699 HQMPHPGVTAMDMASTACAEASRRSRSPPTSPITGSTVEPYRRYVSSKYSQAGPAST 6758  
Qy 220 ----- 219  
Db 6759 RRRPOACANAAASAPLCGETRPMKSRKSSLATWSAADRXRAGSTGIGITGASSPPQWA 6818  
Qy 220 ----- 219  
Db 6819 RCARELXVQPTGLTGRVVKWCELMKAVPGSGPAGPMTPTPLWSWTTTSTSAVRSRHSISSGS 6878  
Qy 220 -----THSF 223  
Db 6879 PAGTTVSPGRPAGLPAGPSRKTSPWSSSSPSIRLATHTHMPYPYCGGTETIOASLPGTRSG 6938  
Qy 224 YLVNAMSRLPRVPKYINGTKLKNTRKLRKQA----- 257  
Db 6939 MLTSATFNGLIRTPP-----SPRRARRARAAGPGGTAWSPAGRRASAAPRWSP 6989

QY	258	-----	257
Db	6990	ALRSRPGRRRCPXGPPRASAARRAGRPSPGASRRRXXRGSSGCAACGPGCTCWRPSR	7049
QY	258	-----	257
Db	7050	RRPGACGPGCAGPPASPSPGPGGTGSRPSRSCAAPRAARSAASRSGRRAPNPAHRA	7109
QY	258	-----	257
Db	7110	PASVGPAPSAAPGCRDGRSRRRRRAARPATGPCASTRXTRRAPWSRSGSTAAGRRCRT	7169
QY	258	-----PVK-----	260
Db	7170	PAIPRFVVRVAPGFAAASRCRGXGTRSPRRTFGRGCRRCSPRRSPDCRCGXSPRAR	7229
QY	261	-----	260
Db	7230	SRAHARHCRPPSRCRSRPPGRSRWCPCPARSDTGPASTARWSRXXSRPAADARNSS	7289
QY	261	-----	260
Db	7290	LPFESGPRRAASLQDGTDRVQLGRVQRQLVEHERVGGDDRAVRPGHPGPGRIELLRR	7349
QY	261	-----	260
Db	7350	GGQHLAAEPAHPVGVIDDDQPSGLLRRRDLPLVHRVDRARIHHLDRLLVRHGLRGQGD	7409
QY	261	-----	260
Db	7410	MHALAGDDGDLARAVDPGRPERHRAELAGDLLDRAHQARREDGHRVVRTDRREQAV	7469
QY	261	-----	260
Db	7470	GVRGGGQDLADADCEQAVGMVGLAGHARARTGARHQIGDRDLHPTGHQRELYLVG	7529
QY	261	-----EQFEK-----	266
Db	7530	DLLEHQTEQRGDLVLHRSLSLPHRRRAGQAGELLRQGGVDDASVAEALVQALGDLGAD	7589
QY	267	-----	266
Db	7590	PHVLAEDADAVARHLVVQRLVORGVRKCSHQIITFGRGVFSAXRWCASAGSARPA	7649
QY	267	-----	266
Db	7650	RANSTASSISVLVSAMAPAVAPSTRVLWVRSGSRARHSPTSASRXFSPTCSSKEWPL	7709
QY	267	-----	266
Db	7710	NRXVRQSRKVGPPRRQRSTASAAAASATASASIPSTARAVRPNPAALSTGSPASARSGR	7769
QY	267	-----AKTQSTTTPFS-----	279
Db	7770	LVAXPLFSQNTGSRHRAARLNASSTPWLSPAPSKKATATRPSRRWNDQAAPAMGA	7829
QY	280	-----YTTGAA-----LNVTI	290
Db	7830	PAPTMPLAPXTPLTSAACIEPPRRPRPAVPVPHSSASIRFGSAPLARTPWPRWLVLT	7889
QY	291	NTVYSIT-----	297
Db	7890	SSSVSAAQTAAATASWPMXGMPPTTIPSDDIARQSSNLIRRMVXRKXTARGSALIAG	7949
QY	298	-----TAARV-----	303
Db	7950	FPSSVVGXASGAPTAAARRAPAAVVARRRRPARSRPRXPDCRTAPGRTPMR	8009
QY	304	-----	303
Db	8010	TAGSTPRWCPASASSPARRRCPASSRAAAAHTPRRPPGAAAAAPGRSRPRXASSATAA	8069
QY	304	-----	303
Db	8070	AARAPPAPCRGPSPASASRPAPVRSRGGXSARSASXTCRRSCGPARRPTAPSASGTSRSP	8129
QY	304	-----	303
Db	8130	GRPRWXPACWXRHTSSPVRTAPASGTRGSHRRPPRTWRCRTTRTRPPSPXSGTSCSRRS	8189
QY	304	-----STST-----	307
Db	8190	ARDCAASXSWRGPRPPSRXRPSPPPARPSSDRLLPSTGTVGRVQEHVIQRRAAQRHID	8249
QY	308	-----IAYR-----	311
Db	8250	HAVALPAQQLHDLHEVADPVAYRRVDTAGLVHGLGPREAGDLSRGGRQVLGGRQGHHD	8309
QY	312	-----PD-----	313
Db	8310	PVAHQLOGLGGLDDPAARVDHRDLVGEFVGLLQVLRGEHQHGALLDHRTHDLPDLVA	8369
QY	314	-----	313
Db	8370	VARVEAGGLVQEBEHLHGEEGGQVEPAPHTARVVLEPLVGLGQVQLLQGFERTALRL	8429
QY	314	-----	313
Db	8430	PAHVVEAAPHLQVLAAGQDLLDRRVLAGDADVPAHLGGLAGHIEAGHQGTAAVRLEQGA	8489
QY	314	-----	313
Db	8490	QDAYGGGLARAVTEQCAHGALGHDEVEPVQRRLGSLVGLAQPLGNOGSPRRPAHTPPSAI	8549
QY	314	-----SSFMKSIMA-----	322
Db	8550	SSDGSAAFMNALSARRWAAGLLQSRPSHHSASSASTVARRFSSSGLRATDRSSGHCTTL	8609
QY	323	-----	322
Db	8610	ARSASRPSRAKSSRSRSPHRFRCGATWSHRFTSVLARRTFATKTSHFAGWMRMSKTSP	8669
QY	323	-----	322
Db	8670	TPSQDAPITLLKCGKRKTFSGSTVRLXLXTMSTFSAFPVSISSRSIAKWPCKMGKSP	8729
QY	323	-----	322
Db	8730	AKYSRMRWAFSRVAHSSSSRYGSIITADDFHGRFLAPOELTGXAHITHRRLQVLNLR	8789
QY	323	-----	322
Db	8790	MSRKKGNSNBPFGAVRSARSASAKTSVANRSLMLCRXDVAPRSXXQXSHADTRSPASI	8849
QY	323	-----	323
Db	8850	ASRVRMLGLLNASVPLTSSGLSCAXRELGAQPTASALS�KTGSASDSYSSLWYIESST	8909
QY	324	QLR-----	326
Db	8910	QORPXTPSMEAHRWTHGSTMVGLGGGWIISRPSPASPARAACWRSAGCGPRTARPGPRR	8969
QY	327	-----	326
Db	8970	GLGRRRPGRRRGAGAGAGRPGPSAAAARPPRPPASRAPGRGSAAGAADRVRXAXP	9029
QY	327	-----	326
Db	9030	FGPRXTRPARPGPRVRRRSSATGPARRAPRRHAWTAPACRPAAAAGAPRADAGPGPSRV	9089
QY	327	-----	326
Db	9090	PGPWRAHRTSRPAGRTARPRPARPPSGPRRRASRSPAGRRGAGRTARCPLKRSAS	9149
QY	327	-----	326

Db	9150	APTERRAARPARCGGRADGRRPRRRGADRSGAPSGARXXPPCRMARRPGAPAGRARPAAA	9209
Qy	327	-----	326
Db	9210	GGGAGSPPPVLLVREPOAGDRGRLVPPEVLLHRAVLGLLVPVLPGRGPLLQARAQHLLL	9269
Qy	327	-----	326
Db	9270	EPHVLGHVRAELTLHLAGPVLVDDPALLAHHRRGAVQCQDVVQGVARPHHGDDGGRV	9329
Qy	327	-----	326
Db	9330	PLGGVEPVGLGDLVLVPPVGLRDPDVTVLGRVVEVDHVVDDATAAGDGGIGVHLAVT	9389
Qy	327	-----	326
Db	9390	VPQPGRARGLEHHAIVPGDEVRPDQGLHGAQQQMARQPGHHRGLVHDVLDADRGRVG	9449
Qy	327	-----	326
Db	9450	LVGLQRPDGVVHGFAGPVVDQAGRLLHRLQFGGAQAAHQHTLPPEGGDVLRQAGL	9509
Qy	327	-DLA-  :	329
Db	9510	PDMAFGDREGGEGGCGIHVECTSPRGERADRGRSATKYHCVRSSAARAARSPPGSA	9569
Qy	330	-----	329
Db	9570	SQGRVVSXVPRSTSSAHRSSLLVPSXSIWSPPCSKTTGDCRRSSAAPRAPTSPP	9629
Qy	330	-----	329
Db	9630	STSTLTRSTRASSGTTSSRRTVSTAISSRVSSARSYGRRRRPLXPGLETTWTXLSRPGVA	9689
Qy	330	-----	329
Db	9690	AAAACTVTSGKRRSCSAYCGSGSTAKCRPGAARMCMVRISSPPAPMSRQPSGAQTR	9749
Qy	330	-----TWVYTL-  :	336
Db	9750	SMISTVNSYSQSLSWGTSVATSDRPLFARGRLVPAAVQLGRVQQQFVQGAQLGLVEVE	9809
Qy	337	-----	336
Db	9810	VHGPDAVLQVLKLGADDDSGVGQHPGKRELGTVPFAGGELLQPVHGREQLGEAFA	9869
Qy	337	-----RYQNPF-  :	343
Db	9870	LEGRGVGAPVVGREHLAGTSSRQHPVRERRPPHDADPGSGAVRQHLLLDAAVEHYERHL	9929
Qy	344	-----	343
Db	9930	GDLRGDDGGEFGHFLDGEVDAHVPGEPFLHEPRQTRRLGRDRDPRVGPVDQEHVHVGP	9989
Qy	344	-----	343
Db	9990	QRAQAVQVGAQBTGVAVADHTVAPQLQAALRDDDVGTGAQVQAQVQAQSLRGHARAV	10049
Qy	344	-----	343
Db	10050	GVRGVEVDAQVERAPDLRDLGRLPAAPEPAQRPGAQDGDGLGALPSQRAVPHAASCSL	10109
Qy	344	-----CEP-  :	346
Db	10110	RRVRKQTPPWSGCDPTGTVSPASAASLRTTSSVPSQDNPLTATGCAPGXLLITYSWNCGLV	10169
Qy	347	-----GRNR-  :	350
Db	10170	TSPKSLKXNRPPGRSTRRISSRXRTSKAFRWMKTLSPANPASKEASGSPARKSFRMMW	10229
Qy	351	-----	350
Db	10230	XVTLGRPAQVARVFSRNRDLRTSAHSLRPLTRGASSRAMTSPPLISNRSVSGVRDSAAANW	10289
Qy	351	-----	350
Db	10290	SIRSARRRYSRGXXYPASRTCMPGSSSHSQTTLSSTLRGIRTPLPAGVWVWPLHSXC	10349
Qy	351	-----	350
Db	10350	RSNRRRRGPCPAGRCSPTRAGCAXALRRAAPRASSPASSWRSCRGSAGRARPPPARGT	10409
Qy	351	-----	350
Db	10410	RSAAPLRVRRSPRSDGRSSHRRCXFPQGRTPSPGSPVAPCPRRSTATSGRRCRX	10469
Qy	351	-----	350
Db	10470	WMAAVRRAFCRXESAPPARRRGASAVRRSXTGGQCRCPRPACARRRRSAPRXTVWPRTPT	10529
Qy	351	-----	350
Db	10530	NAGRCXRCFSSSRTAGARRDPRRSPRWRPRGRRAPPPAAARSRSRPSRPAASPAAP	10589
Qy	351	-----	350
Db	10590	DSFGGRPATPGRRRGPAFRRTARSRRPAAAXTAPRRARRGRSCRWTGCPRRSPTDP	10649
Qy	351	-----	350
Db	10650	AXXAVTGGSRXSRRHPXRRPAPVPSSARLSRAPPSLGGVVHDLGDVLRQPAVRHEAGQQ	10709
Qy	351	-----TAVSBFMK-  :	358
Db	10710	PGDRGRRCPDLDAVPOFLRRPAAVVEQHPAQRLDLLGRGLKAGQOGGAGGIDTAAAD	10769
Qy	359	-----	358
Db	10770	QRVVVEGQLRNPRLDHAHVAVGAHVHQVAVREDELQRAVAQHPDAGRQSRQLHLF	10829
Qy	359	-----	358
Db	10830	LRRREHLLRALPHRRVDGVPQRRYVGAEDARGVAPDDAHASAPDRPQLPRVPALGAVVG	10889
Qy	359	-----	358
Db	10890	HVAHQPOQVQAQRHHRVXKGVHRVLDPVAAPLDAVLLEARQVRVDQVEPRGLHGAD	10949
Qy	359	-----	358
Db	10950	PGRHIDARHAPERDQCALDGHVVHHQVGPVVAEAHRVWIAALHRVDRLPEVAQVAG	11009
Qy	359	-----	358
Db	11010	VEVVERLHRLPGRVQRQRPRTGLRRCVHGLHIERQPGAFHVRVVEHRDRHQAGRGDPA	11069
Qy	359	-----NTHVLIR-  :	365
Db	11070	FLRRRARHDLVSRGAQPCQGHHAHVMRSPVRGGEHNPBKDSLTCRSMASWRSAAQRS	11129
Qy	366	-----	365
Db	11130	GRRRCRVRGLDDGPPPTAVGVARTVDEPDVVHHHVALADRNRHSGRAHLRHLGVV	11189
Qy	366	-----	365
Db	11190	GEQVRSRNLQASVVRSGVVQADAQIADRADVDVAVTVPWHLRGPVRLLEDERVCGGHD	11249
Qy	366	-----	365
Db	11250	RAVHERGBELRQFLAVQQVPRGPVLLQVHVAVEDGTILALDGEVPPVVVGAQALVGEAG	11309
Qy	366	-----	365
Db	11310	DEFFGARLETGDRVRGENRARVREAVPSEPLDRRGVQCRRGHGCVLLSPRGSVTSGAR	11369

QY	366	-----	365
Db	11370	RPCRRRCRRPLPWPAPAAAPRHRXSPAPGRASATAGPASARRXGRGFGARGLGRG	11429
QY	366	-----	365
Db	11430	GPRTEVGHSVRAAPVVGGEVAVGDGADVGEQLPGPLHVVEPFAERGVAVEAPGELHR	11489
QY	366	-----	365
Db	11490	LHVPEPEHVGDAVHLHDELGAQVLFALQGEDLDAGLSHRQGVRAQLQAELODALGP	11549
QY	366	-----	365
Db	11550	RERREVRGFLGLRLAQOVVARGGGVDPVHVHKVHPRLFTYTPNLKIVHCCLTSTV	11609
QY	366	-----	365
Db	11610	RCASKRDGLGWSPVGNAPQVGHITLRLDALSDLDLEGAQFGGADGGGPRGGPGEV	11669
QY	366	-----	365
Db	11670	RQGVRFVAEDLAGLLGGVGDHHLVPGQVQVAGDTGRTTHGLQLQHGVGAGRADV	11729
QY	366	-----	365
Db	11730	EGPAPGLLLRGGAHDDVHDVLDVAERPHLTAVAVDRDLTAODLVEBNGDRIAIGVAHVL	11789
QY	366	-----	365
Db	11790	PFAVHVHTECGEQALGPHRRFQALDREFRVAVTVGYRDRRLPHRGLAVPVHRDGR	11849
QY	366	-----	365
Db	11850	ETEERDAGSORRYDEVHAAHHVCAVVEALHVRTDLRRERREVEDVESLGEPEVDQLH	11909
QY	366	-----	365
Db	11910	IHRPGEQVRALEDTLGAAAEVVEDRYVPAAEQSLRHMRADAGSAGHTVVRHQISWK	11969
QY	366	-----	365
Db	11970	IWYTTNKEFGQRRGSPRAGSPVLMGAVGTARAMPDSARSFNGSHQSRFSAYQRTVSAR	12029
QY	366	---NET---	368
Db	12030	PCSNDRTGAXPSAPRASHSEXRRCPFRSGTSTRSQPAPRASSCCVSSRLVSSAEP	12089
QY	369	-----	368
Db	12090	MXTCPSPRSATSRMPRTWSDTCTQSRTLPPSYGMLWPSRVVNSGTTFSGNKXGP	12149
QY	369	-----	373
Db	12150	XLLPHRVTLTGSWVRAXARATWSPALDAAAGELGAIGASSOEVPVMLPYTSSVETCT	12209
QY	374	-----	381
Db	12210	ILAVPASRHTCMSAWTPTCLVRMNGAASAIIDRSTCDSAAKLTTTSWPGTTSSASSLASQMS	12269
QY	382	-----	381
Db	12270	PCTKYRPATGSFARFAPAXVLSSTVTRAGSPRTSSRTYWEPMKPAAPPMTTRCPRSA	12329
QY	382	-----	381
Db	12330	PALPSRMFPXAAFTVASPASAGLFNRRRPGCGLPAPAGARSTSGGRRSPRSPGRSPSR	12389
QY	382	-----	381
Db	12390	HGRLPRSKHAARSSTGSCARPSCPRWEXRDRGRSPAPPCPRXRDRDRRCGSRSSPP	12449
QY	382	-----	381

---

Db	12450	PPRLGPPVVRWPRPPRRRCGSPCTPAGASSPQRRXRRSPSTGSRSPWRPRARGRPAT	12509
QY	382	-----	381
Db	12510	AGRRXSPARSARTAGCPRPRAPRSXRRTERRSRGRSWGPGASPRPSRGRSQVRPRRT	12569
QY	382	-----	381
Db	12570	CAGRRPGRAXPARRACRGGSPRRCRPRRSCRGTXRSRRRRRSGHGTARRRRPARAGR	12629
QY	382	-----	381
Db	12630	PACRXRPGRTPRRLPRKPGSPRPSRPSRPPIFRPPXWNPVHRREQVWPGPVVWPR	12689
QY	382	-----	381
Db	12690	YAXSAAASQPRPIIXAPSVYPTCGORRQIRSFQGNRWHMSPNNDRGTGVPFRSLGF	12749
QY	382	-----	381
Db	12750	RAGWRGFTPPSPADMSYGVSAFRTHGRPPSGVGVSGDRRCRXXWMAASRRCWPARG	12809
QY	382	-----	381
Db	12810	RGASPAGGPPSSPSRAPAPAGGERAGFWGRVAPARSARAPRAPGNAAPATDPT	12869
QY	382	-----	381
Db	12870	TPRPRRRPARCPCPARPAARPGSAGRAGSPGVPPRXPYPCARCCRRGDRARPPSP	12929
QY	382	-----	381
Db	12930	PVACRRXSRRRARPCCAGGPRRRPARSTRRPAAGPRTRRRARAPERAVHRDRRTVAR	12989
QY	382	-----	381
Db	12990	RSNRNRSVRPPRPAAGRPGRCRVPVAGARPGPAATAPDPRRGPAPSPGRHGRARRSG	13049
QY	382	-----	381
Db	13050	RRRPRSGRPAAPPRRGRPGSADPPSAPGSSSCPGGEGALRPRRRPGTYPPARRPG	13109
QY	382	-----	381
Db	13110	RSRPRXGRGCPGTGGCRRPACCATAAGRCPRRRRACPAACGSRRRGRARPLAPRRWNATA	13169
QY	382	-----	381
Db	13170	GGAARGAAGLRGRSAGAPTSPRRRSGRRRPRRRRSGRAGCAVASVRRTRPPRRVR	13229
QY	382	-----	381
Db	13230	PASRGSSARPGAGTARTFGGRPRPRXSRRRAAPGCAGSPSTRRRSWPAGPRAPS	13289
QY	382	-----	381
Db	13290	CRAACAVRWSARRRPRAAAGRAGGCGCGRPRRRRPAVRPAPRRSSGRRPSRPRGHRR	13349
QY	382	-----	381
Db	13350	SRWTSSTRWPRPRHCPRAARRPRRADXPSPSRHAPRPAATAATYGRSAGPRRSRAP	13409
QY	382	-----	381
Db	13410	RXRRRRRSHRPSGRXRRRARRPVTATARRAASRRSWSGRRRRCRGRPPAASAGARTRS	13469
QY	382	-----	381
Db	13470	PXYTPAVRRSPRRRPARLPVAVGSPRPTGTRDPTPRDRGRHPAARPRXRDRARVPEWP	13529
QY	382	-----	381

Db	13530	ARAAARRVRRGRSRRPPGGPAGTADGPGCARPPGQRPQRPARRARPRGHGAPPGRPRT	13589
Qy	382	-----	381
Db	13590	PGTGFRPAPARGGARWAPVRGPHGPRCRRNRTPQPPGGADARATAGPGSPAGCGRP	13649
Qy	382	-----	381
Db	13650	SRSGSGRRPAGWPGSGPRGRGSPXXPRPRRPPPPDARCWTSPTRPTAPDRSGHVRR	13709
Qy	382	-----	381
Db	13710	RRSRSPWRRGGYRCGRPRRSQRCQRRPRPSGRASXRAPGRRPXCRCWCARPXTR	13769
Qy	382	-----	381
Db	13770	CRAAPAPGHRVRRPSAAXXPARRSPRGRSRRPTQRSSVRPTRTACSGRRAPARRP	13829
Qy	382	-----	381
Db	13830	CPGRSSGVTTPAGXRRPRAPCRTTAAAGSARRCAAPGRPSTRCRWGRMARAGRARERG	13889
Qy	382	-----	381
Db	13890	WTRTTPRRSPSMRCSAGLPVVRHSHRRCRRTRHRCRSLPGCAAGCPAPVPARPPA	13949
Qy	382	-----	381
Db	13950	SGAAAGSARPARACRTRRRNPLCRRPCNCRSESPSAAPHRTADRSPGTLPSGRGASR	14009
Qy	382	-----	381
Db	14010	RRCCVPSTGPRIRASXHHRTGIRSRPPPERGSPPGQGISRAYSVRCRYSFVPV	14069
Qy	382	-----	381
Db	14070	RRFENLKGAGHHQARGGELQDVVAGGELTGYDVVFHQQMVGAFSGRSPGRIPTWTG	14129
Qy	382	-----YNE-----	385
Db	14130	TAAPRRPPSTARLRLVIVHEGDAGDVRRHPDAGPGQALGHVHVHVLNLGHRNEGGAAG	14189
Qy	386	-----	385
Db	14190	ELLADTGAEVGTAPGDARREORLVTAVORLRLQPLGHRHAFPGEGAPLAAPGRERIAA	14249
Qy	386	-----	385
Db	14250	DEAGAHPPPVDQRRSVVVGDDQAGAVRIGLQVADGEFORAEHAVAALPDLOGEHVPAH	14309
Qy	386	-----	385
Db	14310	QVPLQVCLGRPAVRALVPGARIOQIIEFRRSAPPLHLLRRPQGHFRRGDVAGYGDPAE	14369
Qy	386	-----	385
Db	14370	LVCLGLFGRIEAAFGFPDATHRRYEDAESTDVQVGYLHQPTSLFSGLLDPARTDLHSFD	14429
Qy	386	-----TMP-----	388
Db	14430	DHSQRFVILLEGITXTRSGISRAEFAAPSANDGTRFCTGGRGARLGAVEQVBRRHGDG	14489
Qy	389	-----	388
Db	14490	GAQGDHGHGEQRCQDAEPEPDGEDHQFCQSAVHEHADAQGVDAHLHRHQRCCRRAAQ	14549
Qy	389	-----	388
Db	14550	FADDGHQGDAGPHPGQAAAPAEQLQVGLQADRGEERRQYGGDEFLQFAAGVQGRGL	14609
Qy	389	-----	388
Db	14610	RADGAEQEAGDEMDAYELGGDRQQDAGDDQCEHAAVGQALRPLQDASQGGADEHHHD	14669
Qy	389	-----	388
Db	14670	AEEDRGERGAHQAGRTAAQOQDETREQPAGEIGGGGPHDGVGARPGLEHFAHQOQR	14729
Qy	389	-----	388
Db	14730	QHGEGHAHQDAQEBEGEROEADAVRGVAGVQVRSQSTAEGERHADADSHGPGCAARAFM	14789
Qy	389	-----	388
Db	14790	VGVTTELHQAQEHQROPDLGDETEVGVGEGEGEVVDPSDEAHHRAENDPGDHLAHHA	14849
Qy	389	-----	388
Db	14850	WLAQVPADQRSAGPRDGHGHEGEOQPPRLDMRRIIPRRGTPGAGSRVVLARGAVPOOQG	14909
Qy	389	-----	388
Db	14910	RQQRQHRQOAHGKHAGGAARCHLRHPITQLPAWAAERPGCLRRPGSRVPRPGRST	14969
Qy	389	-----	388
Db	14970	GGPGTGAQAPGPPRRRRSGCCSCACSRCTCPRSPRCRAGASRPGRSTAGXRCPSRRR	15029
Qy	389	-----	388
Db	15030	SQCAAPCPRXGIPGVGHGTPRSAGTRPRRAGPSLRCCGPAPARXPCERSRGGPARA	15089
Qy	389	-----	388
Db	15090	RGFRRRRSARHPPPSGRCARPSTRSGAARSSPGARGRRRRRRRGASRADSSARSPCR	15149
Qy	389	-----	388
Db	15150	RPSAGRPPAXTFWDAGRTSRXRPWRGARRAGMRRSRGSARRPAGTRRRRRRG	15209
Qy	389	-----	388
Db	15210	PLRPGARRRPPAGRCRCRCPRPGRGRGPPGXWRRPCPCAGFPWMSGXWTGSRSP	15269
Qy	389	-----	388
Db	15270	RRTXGSRRCGAAAAPVRNRASAPAPABERTARPRRGSAGTRRXGPVARGARARTRCCR	15329
Qy	389	-----	388
Db	15330	RSNGRXRSSDRSCSXFSSQCSATGPEPADDICGVRPVPFADQGLVVLVLAGEDGTRTD	15389
Qy	389	-----	388
Db	15390	RVPDLPQADRRPRPHLPQPLGVEVGHGPAAPLDRREALARLAVRVDGILLFEDVDENALG	15449
Qy	389	-----	388
Db	15450	VAVAVHAQADQMARAEILLGLPVPVGVGVQLGARGDGLDEAVPLPVAKERPGLVDR	15509
Qy	389	-----	388
Db	15510	LAADVVLPGPVTGGAGRVQHVGAATACLLVVGTRHNGRGTCQCPGDGRADBPFGA	15569
Qy	389	-----	388
Db	15570	DHGDPHGRSSVCGSGSTXTSGSHSSQGLVSMVMNVDEMAWTECMTPAGVLVQTPREIW	15629
Qy	389	-----	388
Db	15630	MNSSLQSSASSLVISTPEPSTNSRYSTKCGXCRPRTPFGYRCSILNSRMGSGTQVK	15689
Qy	389	-----VENKT-----ASDSN-----	398
Db	15690	IPRPSRTVSRLOVQGSTSIFMPSPLVPRHRFAGVEDKTGTCKEISSPVGGAHDSGVR	15749



QY	399	-----	398
Db	15750	QHRVTSREFSGRPDACYPRFRGRNGHATADAAVTPRSATTPEFSQPLADHRDREVSAG	15809
QY	399	-----KTT-----	401
Db	15810	RRPPKAPFARTTQCPRATGTATKYKHYCWXTSVGYKVSRRCLVPPRRSRHLSALFPQG	15869
QY	402	-----	401
Db	15870	VPGENRPHDLIQGRQDHRTHRRXAHHLHQALPOPPGGRVPRQDDGALPGAQPHRP	15929
QY	402	-----PTSPS-----	406
Db	15930	RRQGLRAGNPEIRGVAQPSAGTRQLSHCHGRXAPSPSPWTKRACQCSLSKRVFIYP	15989
QY	407	-----	406
Db	15990	ATIRXLGAHGRPTKAGDAPFRFRAGRPPGCRILRGASPAFCRGAGRTSRVPSGVIAVH	16049
QY	407	-----MG-----	408
Db	16050	TPGTGRLIRLAHASAGPAHDSRRHQAPTRPXMRXDDQDEXCPPIRMGGIALTSXFS	16109
QY	409	-----	408
Db	16110	GGALRLDTRCVGXPLRACKDHTLIRGPDPAIFELRFQXFWRRSISPRGPVPLAGRAQPG	16169
QY	409	-----	408
Db	16170	TFTLRPAAGLPAEPRLLPQDORRAARGHRAHQSGYSGDTEARGRIPSSGRTGCLRP	16229
QY	409	-----FORTFIDP-----	416
Db	16230	RASRTHRCRGIRPSTFYVGFDPQLQCLAGRVRESGPHQVYHQLHELPLESGRQVVPLL	16289
QY	417	-----	416
Db	16290	VEQIRIGVRLDVQVGFAGAGARAVDVVDLVAVRAQQIDHILGLLPDLROGQVAVSFQ	16349
QY	417	-----	416
Db	16350	RVVPGRVVVRDDQORGRVSLPDDPQDQVVRPVGELVQGTRAIDVHADHARDQIIGQVA	16409
QY	417	-----	416
Db	16410	EQGREERLEVAARPDRQGRQVQPEFGDGRGPVPGGRQPSALRLGPGVVDPAGRGVNP	16469
QY	417	-----	416
Db	16470	AHLGACRXLHAVQDELVOAGQPOLHKAVAAARQCPDAPVDLPVGHQGEAGDRLAVHSYV	16529
QY	417	-----	416
Db	16530	DVRLGRARHLGRTGRVPECEPLGVGPESDPHRRHRGASHAALGKBQVFFLKRYSPGR	16589
QY	417	-----	416
Db	16590	YSTVDLLEGDGAALGOHRVIRPNEAPCFQHPSPVQSFVXADASRLTAPMRSEPPILXHS	16649
QY	417	-----	416
Db	16650	XKAVRTMDDXHSRVPPPABRSHTGVTVDKKTADRTNTPGRRPMVAVCAPLENVRHL	16709
QY	417	-----	416
Db	16710	RHNDQNGDEYVLLDRPVTPEABKVATQVNEAGTMRILGATRGLSASRNTVLRWAD	16769
QY	417	-----	416
Db	16770	RHVLFDVDDVRLASAVDAVRAAFRAGAHVVGARLPPPRELRLPWFSLSSGQPHLVGWHR	16829
QY	417	-----	416
Db	16830	DRGDIKIWACMGVDADFAARQGLTFDLDSRTGVNLQSGEDTSFTALMKEAGARELPAA	16889
QY	417	-----	416
Db	16890	RARGRRCRPRPAHPPLPPAPGILLACGVRRRDGTSRRDSARSXPCTAPRRSRAAHCCS	16949
QY	417	-----	416
Db	16950	ACTGPPPSASATDSCCASCASCLLPDGRGTGDPGVGGDAFLQDRGRGPAQIGDTS	17009
QY	417	-----	416
Db	17010	VOGAVLDLARPAGRETGRQVAPGGGGVVEDLQHRAPLARVCRRRRADGAVRRDERRGD	17069
QY	417	-----	416
Db	17070	VADVHVVTGLRTVAEHRGHSALGGQPAVDGDDDLTARALPGPVHVAEPQHAVLGSVQPV	17129
QY	417	-----	416
Db	17130	EQADVLLRAVLGRAVGRPRPGQRVLRVRDRNAPFVERPAGGGBHDPDPGVRRFQHVDR	17189
QY	417	-----	416
Db	17190	LDVDPVDRGCGHRDPDVLGREMADDVGARRPHGLVQGGIGHVDATDDRTWQTPRVT	17249
QY	417	-----	416
Db	17250	GGQVVOHGDLPGLQBGVDHVRADDEARPSGHQTCGSEXPREHGHGPPFLATTASECNR	17309
QY	417	-----	416
Db	17310	RSGHIRENLRLGDDRDRPASPTTDSPGAGDLCVRRQLGASPAGAAALXSLQCGDG	17369
QY	417	-----	416
Db	17370	GSEVLGSGQVSAYRHAVERIDSSDLACGVVLSHAGYPAPFVRLATEIFORALARLPDG	17429
QY	417	-----LWD-----	419
Db	17430	PVTLMDPCCGSVLLTVILGLLHRRSLRQVIASDVDPAPLEAANKLALLSPAGITARELE	17489
QY	420	-----YLS-----	423
Db	17490	RREQSERFGKPSYLEAAQAARLRRLTAEGGALPCAIRTADVDFPRALSAVLAGSAPDV	17549
QY	424	-----	423
Db	17550	VLADLPYGSARTGKRCPRSRWRALLRSLASALPAHAVIAVTRSRKIPVAPVKALERLK	17609
QY	424	-----	423
Db	17610	IGTRSAVLVRAADVLEAGPAAVPAEFPCQCFQCPCFPSCFSLDSQRQWMLQVCPWE	17669
QY	424	-----	423
Db	17670	CGMAEARVLVAGNLHWQAGFSQTVAAYVEAAGEAGCEVRVSGALSRMDEQIPRHLPEPD	17729
QY	424	-----LLFLDEIRNF-----	433
Db	17730	VRWGTHLVFMFEARQVLSDEQIELACSPDRRLIIDFXRPLGRCHRRRQLRRHPLDRE	17789
QY	434	-----	433
Db	17790	LATALLRAORPRPAAQVDGPLARRGRVLPFRNAPAGTASTGARSAGLRAAVHRQOLGA	17849
QY	434	-----	433
Db	17850	VEAVHRGGRGTQGPAAADSGLRVAVVGRGDLPLRLGHHLGRSGXSGVEVQPPVPGHV	17909
QY	434	-----	433

Db 17910 VTEMGRAVITPVLSPLVADTGLFTARMLETLAGSLFVLPEAEFMSSVYEGDVEALT 17969  
QY 434 -----S 434  
Db 17970 GSDPAATLGRLAGDFERHAEIVGRIQDRARARYNPRLRELLRFFRXSAAVPLRRGRS 18029  
QY 435 LRSP-----TYVN-----442  
Db 18030 RRSRPSASSRSPSGCXPCSRITTYARRCXCARLHSCRSAGSAGXSRSTHRSRARPYP 18089  
QY 443 -----442  
Db 18090 GRSACRRPGPPSPSPRPVTPAGRRAGRTASRCSGTGISGRRRRRAPPSPASARPCHP 18149  
QY 443 -----442  
Db 18150 PSEXCCRSPDWSAGRTVPGSPGPGSGRGRRRASPRRRPRAGRSPSPAPAGPGPS 18209  
QY 443 -----442  
Db 18210 RSAPSAPYEEAPAPTSTAPAVRRGACVRSRAPPARRSGGLPSPGTHRRACACCGRAV 18269  
QY 443 -----442  
Db 18270 PATAGNERCVXSABPPASALHVLIPFEGAYLPULLATKLRAPPXGRPGGCRDVARQC 18329  
QY 443 -----LTPP-----446  
Db 18330 ALRPARSAVRPPACGRAARPLAPRPDTPLRPGSGRPPVARRPPGALLVAPHLWR 18389  
QY 447 -----EHRRV-----452  
Db 18390 HGHPAHDDPHRDPRAARRRRRLIRPRAAGHRSRNPYDHRAAYPQLRPADHR 18449  
QY 453 -----452  
Db 18450 GLRRGSRPTLLPRDRPAGRXRPORQMGAGXPNSAPAHFTEILINDLADHAQVILHSRF 18509  
QY 453 -----452  
Db 18510 WRVDADRTYGISGTGLDWELDWSAPWEHLVEESHTWSLLESEAPVCGSIFVAPTWDRT 18569  
QY 453 -----452  
Db 18570 DLYPERWMPFQHSSESGYGVIERVYVTSKSDRPRCDLKGRLRPGWPHAERFTHL 18629  
QY 453 -----452  
Db 18630 IDDAATITHAEWEALLGYEWSRLTAANWLGVRATFRERIGDLPPASEFCPSGGAHCF 18689  
QY 453 -----452  
Db 18690 TLARFGTHADAEILTGDRQMPIPSNWSALAWTGPYCPIGRXXLSNSQSAGGTKSQGPD 18749  
QY 453 -----452  
Db 18750 SDESDFDLRIQDDVDLRLNVAEHLXQSLRKHVRGAHPSTIPAREHLGSQXWY 18809  
QY 453 -----NLSTNSL-----460  
Db 18810 RAARITISLVTRYTRCSSLIRRDHAPGWYFSGGLPIPVKGSRLTSSMLLWASIL 18869  
QY 461 -----460  
Db 18870 RSFLSQDVXSSQPSKSTSLTRHPNORPNPDRSAPATHXRQGRSRTPXAVDGRSPES 18929  
QY 461 -----460  
Db 18930 AACTSPRAALHNPRPRVRCRAWRSRWRNDRCPXPSAERPCGFARGLSTXDFPFPQ 18989  
QY 461 -----460  
Db 18990 VVNNVPLLTSSGRAKTDGSAFPAPSQHKQQAAPWRSRPLLTCAFPSSVVVELRLHV 19049

QY 461 -----460  
Db 19050 EAELDHVPVLHHIVLALHTLPLRAGLGDGAGLHQVGEADDLRLDEALLEVGVDDAGLG 19109  
QY 461 -----460  
Db 19110 GGGALLDPCAGFLRAGRVGLQAGVEADAGEGGERAGLVADGLEEFEGVLVVVELGEVG 19169  
QY 461 -----460  
Db 19170 LQLGVEEDGLGGDQGALLVLEVLVDQLVLVDVEDVEEGLGGQVQVQVQSLVALGALDDA 19229  
QY 461 -----460  
Db 19230 VGBEGVAPLQDLLGLLDGLHLRREVLLDPRVLEAGQDLLDRLEVREDQLGVDRLDVVLG 19289  
QY 461 -----460  
Db 19290 RDLAVDMDDVLVLEGADDLADRVGLTDVREELVAQALPLGGAHADGADVDEVDGRREDAL 19349  
QY 461 -----460  
Db 19350 GGEDLQGLGEACVGDADHTDVGIDRGERVVGRHQHVVLGQVGOGLRADIQTDDSDRERH 19409  
QY 461 -----460  
Db 19410 VATSRXEDPRAGWAGARSFFDLRCXCRRTGRSTSLRRAGPGGAVSNTWPSAPRVX 19469  
QY 461 -----460  
Db 19470 WAIRHISRPLVQWSHTGRDLPGTKEDTLRRRCRRRRRGVGVSLAAGAAVRGGVKTVC 19529  
QY 461 -----460  
Db 19530 SEAGADSFQAGAVRRAATAVSRAAGAVRAGTAGPFGPPGPRPGPPAPPSPSAALVR 19589  
QY 461 -----460  
Db 19590 PVRPRPRYAGPRPRPRPCAGFTPGXPGWAAACSAAPXCSCSATSTRCCSGRRSRC 19649  
QY 461 -----WWW 463  
Db 19650 TACCPCASCXRRCCGACATWRPRWY 19677  
|:|:

## RESULT 15

ABB58144  
ID ABB58144 standard; protein; 7107 AA.  
XX  
AC ABB58144;  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 1224.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US0009231.  
PR 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.  
DR N-PSDB; ABL02247.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 1224; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 7107 AA;  
  
Query Match 21.5%; Score 526; DB 4; Length 7107;  
Best Local Similarity 3.1%; Pred. No. 0.0017;  
Matches 217; Conservative 104; Mismatches 135; Indels 6550; Gaps 71;  
  
QY 2 GRKEMV----- 8  
DB 25 GRKRLVICKVTDAGQFKCTTNAUTTESEIIINYQNRNFKKLKOTEAREKILIDIE 84  
QY 9 ----- 8  
DB 85 LQDTAPCDWKFNGEPIVPSIEIKNMGGKHQLIFSSLDMSNEGEITCESQLSKCK 144  
QY 9 -----RVPRKF-----VLISIF----- 22  
DB 145 LSIRKGESRNIDCPDKFGSIPAPVLLVFPFKVSGTKQTPVBAKLFKDGKPLPVKDVEV 204  
QY 23 ----- 22  
DB 205 AVTDDKVTFKKPSRDLSPYQIKISNGGGETKDQVQICQDVPPQPDVDITDVQTS 264  
QY 23 LVSF----- 27  
DB 265 CVVSFPNPPDDGTPITTKVIERQDLKKGWESVAEVLPSBCLKKIDDLIPKKQYRPR 324  
QY 28 ---INC----- 30  
DB 325 IRAVNAIGOSDPATFRNTILAKDPWDEPGKPAVDLTWDKDHADLKWEAPETDGGDPIT 384  
QY 31 -----KV 32  
DB 385 AYIVEYKEKFSNDWSGVDGDTATVDGLKEGQQYEFRRVAVNRAGPGEPSDKTSI 444  
QY 33 MSKALYNRP---RGL----- 45  
DB 445 IAKCRFVKPPIVGEGLKNVTVKKGQTRFDIKYDGEPEPAATVWKGTDNLKFDNQICLD 504  
QY 46 -----VLSKIGYKL----- 55  
DB 505 QLERNSSITTKSVKRDYTKYKLVLSNSGCTIBSEAQVVLDRPLPPGPFPEETRASH 564  
QY 56 ----- 55  
DB 565 IKMKWRPDDGGCEISGVALERMDEETGRWIPAGEVGNETSDFPKGLTPNKKYFRVK 624  
QY 56 ----- 55  
DB 625 AINKEGESEPLETFDAIVARNPYDPPSPQVDDVDNKSLLKWKRPSPDGRPIITHY 684  
QY 56 ----- 55

DB 685 IVEIKDKFAPSWSEVAKTDDPNECNVEGLKEKNVYQFRVAVNRVAVNRKAGSPSPQTDNHL 744  
QY 56 -----DQLKLE-- 61  
DB 745 KHKNLKQIDRSFTFRVTIKSGRTHKWSVDVLGEPPELHWSMRDDIPLTNGDRIKENV 804  
QY 62 -----ILR----- 64  
DB 805 DYHTDFSITNVLKDSGFYTLKAENRNGIDRETVELVVLGKSPSPKGLAVSDVTASGCK 864  
QY 65 ----- 64  
DB 865 LQWKPEDDGGVPIKEYVVEKMDTATGKWRVGRSGEKEPPSPFDVTGLSLGSEYMRVS 924  
QY 65 ----- 64  
DB 925 AVNEEGESPLTLVLGVAVKADPEPNKPTPEVTDYDQNSISLKWAAPNNDGGAPIOKY 984  
QY 65 -----QLETTIS-----T 72  
DB 985 IIEKKNKTEWEKALEIFPDQLEATVAGLQEGYQFRVIAVNRKAGLSPSPDASVQIV 1044  
QY 73 KY-----NVSKOP-----VK 82  
DB 1045 KYKLLPRIDRSNLKLLIRAGKPIRYDVNRGEPAPVITWYQNDKELKPEELPSSSEIK 1104  
QY 83 NLTMNTEFP-----QYILA----- 97  
DB 1105 NIPYNTKISIIETVRKHTGIYKIIAVNEHQDEATVEVNILAPPSKPRGLDVKDVTKDS 1164  
QY 98 ----- 97  
DB 1165 CKLKKWKPEDDGGKPI SAYQVEKFKQGRWVPLGRTSANDTEFVKLGLEGHEYOQPRVK 1224  
QY 98 -----GPIQNY 103  
DB 1225 AINEGESPLDSDSIIAKNPYDAASKPGTPNIVDYNHVMVKLKWEPASDGGAPISGY 1284  
QY 104 SI-----TYLMFDFYSTQ----- 116  
DB 1285 IIEKKDKPSIWEILLSTNTSVPEATVEGLVEGNIYQFRVAVNRKAGSPSPDATEPHLA 1344  
QY 117 ----- 116  
DB 1345 KPRNLKPYINRDKMKPIKVRAGQPVKFDVVKGEPAAPSLTWFLKTELTSTGTQVRLNID 1404  
QY 117 -----LRKPAK----- 122  
DB 1405 YNTKLTLLDTRKQSQYKLRANENINGVDEAVVEVILDKPSKPEGPIEVSDIHKEGCKL 1464  
QY 123 -----YVYSQY-----NH----- 130  
DB 1465 KWRKPKDDGGIPITGVVIEKMDTATGKWPAGSVDEPKYDIEIKGLDPNHRYOFRVKAVN 1524  
QY 131 -----TAKT-----ITPRPP----- 140  
DB 1525 EGESEPLETESAITAKNPFDPVSAPGLPELEDWDEHHVKKLKWEPPIRDGSPITNYIE 1584  
QY 141 ----- 140  
DB 1585 VMDKDSGEFVKAETDSPVCKGVVKKLEBQQYKFRVAVNRKAGSPSPQTNHVAKPR 1644  
QY 141 ----- 140  
DB 1645 FLKPHIDRYNLKPIVKTGLSISLDINIRGEPAPKVEWFFNNSSVTSDEHSVKIDNVYN 1704  
QY 141 ----- 140  
DB 1705 TKFFVMRAQRSOSGKYIIKATNEVGDEAELEVTVLGKPKGPKPLQVNDITKHSCKLKW 1764  
QY 141 -----PCGR----- 144  
DB 1765 EKPDGSGSPIDYIEIEKLDPHGTQWLPCGKSTPEBAKVIJLHKGKAYKFRVAVNRKAGE 1824

QY 145 ----- 144  
Db 1825 SEDLETEPIIAKNPYDEPDORPGKPEPTNWDKDFVDLAWPPKNDGGAPIQKYVIQMRDK 1884  
QY 145 -----VPSMTC----- 150  
Db 1885 SGRAWDSATVPDGKCMGTGTGVEGHEYPRIIVAVNKAAGSDPSVSKVIAKPRELKP 1944  
QY 151 -----LSEMLN----- 156  
Db 1945 HIDRKNLOKIMRSGQMLHDALIKAEPPAKVVTYNTKTEIKTSDHIKIENEDYKTTFFIM 2004  
QY 157 -----VSKRNDTG----- 164  
Db 2005 PKVCRADRGIVITAKNDSGSDTVEVEVLCKPSKPKPLAVSNVTAETLHLKWKEPED 2064  
QY 165 ----- 164  
Db 2065 DGGDPIEQYLVERMDTETGRWPVLTTKTPEADVTLGTEGKEYLFRKAVNSEGESEPLV 2124  
QY 165 ----- 164  
Db 2125 TDIPTKAKNPFDAADTPGKQIIVDWSGNHCDLKWRAPEDDGGASITGYIIVERKDPNTGW 2184  
QY 165 ----- 164  
Db 2185 QKALETSTPCKARVNDLIAKNKYQPRIMAVNKAAGSKPSEPSDQMTAKDRFAPPKIDRT 2244  
QY 165 ----- 164  
Db 2245 NIKDITIKAGQHIFDIKVSGEPPATKWVLHNLKARLENDSDNIDMESYRTKLTVPISK 2304  
QY 165 -----EQCG----- 169  
Db 2305 RFHSGKYTLKAENESGEDEASFEVIVLDKPPGPEPLRVTDVHKEGCKLKNWAPLDGGGL 2364  
QY 170 ----- 169  
Db 2365 PIDHYIIEKMDVESGRWLPSPGRKESFAELNNLEPSHEYKFRVLAVNTEGESEPLETGEQS 2424  
QY 170 ----- 169  
Db 2425 VIAKNPDEPGKGTPEAVDWDKHDVLDVWRPPIINDGGSPITGYVVEKREKGTDKWIKGT 2484  
QY 170 ----- 169  
Db 2485 EITIPCLGEBCKATVPTLNENCEYEFVRKAINAAGPEPSDASKPIITKPRKLAPKIDRK 2544  
QY 170 -----NFTFNPMFF-----NVRWNTKLYV----- 190  
Db 2545 NIRTYNFKSGEPIFLDINISGEPAPOVTWNQNNKSVQTTSFSHIENLP-YNTK-YINNNP 2602  
QY 191 -----GPTKV----- 195  
Db 2603 ERKDTGLYKISAHNFYQDQVEFOIINIITKPGKPEGLEVEVHKDGCKLKNWPKKDDGG 2662  
QY 196 -----NVD-----SQTIFYLG 206  
Db 2663 EPVESYLVEKFDPTGTGLVPGRSDGPEYNVDGLVPHGDYKFRKAVNKEGESEPLETLG 2722  
QY 207 ----- 206  
Db 2723 SIIAKDPFSVPTKPGVPEPTDWTANKVELAWPEPASDGGSPIQGYIVEVKDKYSPLEWKA 2782  
QY 207 ----- 206  
Db 2783 LETNSPTTATVQGLIEGHEYQFRVALNKGGLSEPSDPSKIPTAKPRVLAKIDRNL 2842  
QY 207 -----LTALLRYAQRNCT 220  
Db 2843 NITLSSGTALKLDANITGEPAPKVEWKLSNYHLSQGNVTIETPDYVYTKLVIRPTQRSDS 2902

QY 221 HSPYLVNA----- 228  
Db 2903 GE-YLVTATNTSKDSVLVNVVITDKPSPNGPLQISDVHKECHLKWKRPSDDGTPIE 2961  
QY 229 -----MS 230  
Db 2962 YFQIDKLEPETGCWIPSCRSTEPQVDVTGLSPGNEYKFRVSAVNAEGESQPLVGDESIVA 3021  
QY 231 RNLPRYP-----KY----- 239  
Db 3022 RNFDFEPGKPNLKAATWDKHDVLDLAWTPPLIDGGSPISCIYIEKODKYKMERALDVPA 3081  
QY 240 ----- 239  
Db 3082 DQCKATIPDLVEGQTYKFRVSAVNAAGTGEPSDSTPIIAKARNKPPIIDRSSLVEVRK 3141  
QY 240 -----INGTKLKNMTMRKLKRKQ----- 256  
Db 3142 AGQSFTFDCKVSGEPAPQTKWLLKKKEVYSKDNVKTNDVNTKLKVNATRSDSGIYTV 3201  
QY 257 ----- 256  
Db 3202 FAENANGEDSADVKTVIDKPAPPNGPLKVDINSECTLHMNPDDGGQPIDNTYVVRK 3261  
QY 257 -----APV-----KEQFEKKAKTOSTTTPYFYSYTTSAAL--- 286  
Db 3262 LBETTCRWIPAGETDGPVTALKVGGLTPGHKYKFRVAKNRQGTSEP---LTTAQAIIAK 3318  
QY 287 -----NVTT 290  
Db 3319 NPFDVPTGPTIKDFKDFVLEWTRPEADGGSPITGYVVEKRDPSDWEKCAEISD 3378  
QY 291 NVT----- 293  
Db 3379 DITNAHVPDLEGLKYEFVRVANKAGPSPSDATETHVARPKNTPPKIDRNFMSDIKT 3438  
QY 294 -----YSI 296  
Db 3439 AGNVFEFDPVVTGEPPLPSKDWTHEGNMIINTDRVKISNFDRTKIRILDAKESDGTGYTL 3498  
QY 297 TTAARRVS----- 304  
Db 3499 T--ARNINGTDRHNKVTILDAPSVEGPLRNGDVSKNSIVLWRPDKDGGSEITHYV 3556  
QY 305 ----- 304  
Db 3557 EKQDNEAMRWVPVGDCTDEIRADNLIENHDYSFRVAVNKGQSQPLTTSQPIITAKDPY 3616  
QY 305 ----- 304  
Db 3617 SHPDKPGOPQATDWGKHFDVLEWSTPKRDGGAPISSYIIKKRPKGQWERAADVLDGNCK 3676  
QY 305 -----TSTIAYRP-----DSSEFMKSI----- 320  
Db 3677 AHVPELTNGEYEFVRVIANRGGSPDPSSTIICKRFLAPFPDKILLANDITVHAGKR 3736  
QY 321 -----MATQRLDLATWY----- 333  
Db 3737 LGWTLPPIEASPRPLITWLYNGKEIGSNRSGESGLFQNELTPEIVSSLSRSDGRTYLLIKN 3796  
QY 334 ----- 333  
Db 3797 EHGSPASAHATVLDLRPPKGPLDITKITRDGCHLTWNVPDDGGSPILHYIEKMDS 3856  
QY 334 -----TTLRYR-----QNPFCEP--- 346  
Db 3857 RSTWSDAGKSTHIVHDVTLVRKEYLFRKAVNAIGESDPLEAVNTIIAKNEFDEPDAP 3916  
QY 347 -----SRNRVAVSEFM-----KNT----- 360  
Db 3917 GKPIITDWRDHDLDQWAVPKSDGGAPISEYIIQKKEKGSYPTWTVRHPVSNKNTTPIE 3976  
QY 361 ----- 360



QY 431 ----- 430  
Db 6197 NLHDAFEDDEMILILEFLSGGELPERITABGYVMTAEVINYMQICEGIRHHEQNII 6256  
QY 431 ----- 430  
Db 6257 HLDIKPENIMCQTRSRSTNVKIIDFGLATRLDPNEVKITTTGTAEFAAPEIVNREPVGFT 6316  
QY 431 ----- 430  
Db 6317 DMWATGVLSYVLLSGLSPFAGDNDVOTLKNVACDMDFVESFKYISEEAKDFIRKLLVR 6376  
QY 431 ----- 430  
Db 6377 NKEKMTAHECLLHPWLTDGHSAMQEIINRDRVIAYREKLRRKYEDPERFLLPIGRLSEY 6436  
QY 431 ----- 439  
Db 6437 SSLRKLIMEKYKIHDAVDRRAAPRFVIRPSQFCYEGQSVKFCYCRCAIATPTLTWSH 6496  
QY 440 ----- 445  
Db 6497 NNIELQSVKFMKRYVGDDYFIINRVKLDGRGEYIIRAENHYGSRREVVFLNVQLPKX 6556  
QY 446 ----- 450  
Db 6557 QPRYRTESTFVRREPLPYTFWQSESETPSPFTLLRPRVMQARDCKLLCCLSGKPVN 6616  
QY 451 ----- 450  
Db 6617 VRWYKDGRELSKEYAMTHSDGVVTWEIIDCKPSGSKYCKATNCHGTDETCVVIVEG 6676  
QY 451 ----- 456  
Db 6677 EWTPEQAQLAHNPLYSGDRKYIEQPIKPAPLPVITSROYTSSVQNTSEPOGDKVNSN 6736  
QY 457 SNS ----- 459  
Db 6737 SNSSGISNKKYASNSIQAPGSPSRSATKELILPPDDSLMCKPEFTKPLHDLTIHDGE 6796  
QY 460 ----- 459  
Db 6797 QLILTCYVKGDBEQISWSKNGKSLSSDILDLRYKNGIATLTINEVFPDEGVITCTAT 6856  
QY 460 ----- 459  
Db 6857 NSVGAVETCKLTIQPLDKNINKRVNAGDNAPKIVSHLESFRVRDGDVNLACRIIGAQ 6916  
QY 460 ----- 462  
Db 6917 HPDVVLHNNKEIKPSKDPQYTNNEANIYRLQIAEIFPEDGGTYTCEAFNDIGESFSTCTI 6976  
QY 463 ----- 465  
Db 6977 NVTVPGETKQPSFVKPPTSVSVLEGEGETTFCEIDSELINLWLK 7022

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:33:27 ; Search time 40 Seconds  
(without alignments)  
1118.520 Million cell updates/sec

Title: US-09-942-146A-1  
Perfect score: 2451  
Sequence: 1 MGRKEMVVRDVPKMFVLISI.....PEHRAVNLSNSLWNLQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.0  
*Gap extension penalty reduced to zero to allow for deletion.*

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2417	98.6	466	2 S09837	hypothetical prote
2	681	27.8	26926	1 I38344	titin, cardiac mus
3	575	23.5	15281	2 S41309	cyclosporin synth
4	555	22.6	10797	2 T30192	probable peptide s
5	520	21.2	8243	2 T31307	type I fatty acid
6	500	20.4	6658	2 T13931	projectin - fruit
7	500	20.4	9376	2 T14593	syngomycin synth
8	498	20.3	7829	2 T15789	hypothetical prote
9	479	19.5	13055	2 T16580	hypothetical prote
10	473	19.3	7160	2 T27935	hypothetical prote
11	472	19.3	5105	2 T32650	hypothetical prote
12	471	19.2	7962	2 I38346	elastic titin - hu
13	469	19.1	6805	2 S20901	titin - rabbit (fr
14	469	19.1	6831	2 A88852	protein unc-22 (im
15	469	19.1	6839	2 S27242	twitchin [similar
16	465	19.0	4688	2 F82885	hypothetical prote
17	465	19.0	5005	2 F82884	hypothetical prote
18	465	19.0	10223	2 T30225	polyketide synthas
19	458	18.7	6669	2 S55024	nebulin, skeletal
20	452	18.4	5825	2 T12117	polyprotein - fava
21	451	18.4	4924	2 T50176	probable peptide s
22	450	18.4	6359	2 T31679	bacitracin synthet
23	449	18.3	4563	1 LPHUB	apolipoprotein B-1
24	449	18.3	6486	2 T31076	tyrocidine synthet
25	447	18.2	8563	2 T30226	polyketide synthas
26	444	18.1	5147	1 IJFPTM	cadherin-related t
27	443	18.1	4466	1 S17231	dynein beta heavy
28	443	18.1	4466	1 S17653	dynein beta heavy
29	443	18.1	4589	2 T14914	dynein beta heavy

30	443	18.1	4767	2 T31345	hypothetical prote
31	441	18.0	5232	2 A45086	HC-toxin synthetas
32	441	18.0	5369	2 T44807	mycosubtilin synth
33	440	18.0	4644	1 A38905	dynein heavy chain
34	437	17.8	4436	2 E71086	hypothetical prote
35	437	17.8	7576	2 T17428	FK506 polyketide s
36	436	17.8	5255	2 T31677	bacitracin synthet
37	435	17.7	3791	1 YGPLV8	alpha-aminoadipyl-
38	435	17.7	4588	2 T28667	dynein beta heavy
39	435	17.7	4930	2 E69679	polyketide synthet
40	434	17.7	4092	1 S38128	dynein heavy chain
41	434	17.7	4196	2 T43274	dynein heavy chain
42	433	17.7	5138	2 E96695	hypothetical prote
43	432	17.6	4485	2 T08044	dynein gamma heavy
44	431	17.6	4725	1 A44357	dynein heavy chain
45	431	17.6	4910	2 S64942	probable membrane

ALIGNMENTS

RESULT 1

S09837

hypothetical protein UL74 precursor - human cytomegalovirus (strain AD169)  
C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004

C:Accession: S09837

R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horenell, T.; M.; Barrell, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; PMID:90269039; PMID:2161319

A:Accession: S09837

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-466 <CHE>

A:Cross-references: UNIPROT:P16750; EMBL:X17403; NID:G59591; PIDN:CAA35389.1; PID:G17808

A:Note: this sequence was submitted to the EMBL Data Library, December 1989

C:Superfamily: human cytomegalovirus hypothetical protein UL74

C:Keywords: glycoprotein

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-466/Product: hypothetical protein UL74 #status predicted <MAT>

F:75,83,87,103,130,157,162,171,219,242,288,292,350,367,385,392,399,433,443,454/Binding

Query Match 98.6%; Score 2417; DB 2; Length 466;

Best Local Similarity 98.9%; Pred. No. 3.9e-60;

Matches 463; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy	1	MGRKEMVVRDVPKMFVLISIFLLVSPF	INCKVMSKALYNRPWRGLVLSKIGYKLDQLKL	60
Db	1	MGRKEMVVRDVPKMFVLISIFLLVSPF	INCKVMSKALYNRPWRGLVLSKIGYKLDQLKL	60
Qy	61	EILRLQLETTTSTKYNVSKQPVKNLTWN	TRPPQYVILAGPIQNYISITVYLNDFDFTQLRK	119
Db	61	EILRLQLETTTSTKYNVSKQPVKNLTWN	TRPPQYVILAGPIQNYISITVYLNDFDFTQLRK	120
Qy	120	PAKVYVQYVNHAKTITFRPPQGRVPSMT	CLSMNLVSKNDTGEQCGNFTTFNPMFF	179
Db	121	PAKVYVQYVNHAKTITFRPPQGRVPSMT	CLSMNLVSKNDTGEQCGNFTTFNPMFF	180
Qy	180	NVPRWNTKLVGPTKVNVDVSDTIYFLGL	TALLLYAQRCNCTHSPYLVNMSRNLFRVPKY	239
Db	181	NVPRWNTKLVGPTKVNVDVSDTIYFLGL	TALLLYAQRCNCTHSPYLVNMSRNLFRVPKY	240
Qy	240	INGTKLNTWRKLRKQAPVKEQEKAKKTQ	STTTTFFSITTSAAALNTVNTVYSITTA	299
Db	241	INGTKLNTWRKLRKQAPVKEQEKAKKTQ	STTTTFFSITTSAAALNTVNTVYSITTA	300
Qy	300	ARRVSTSTIAYRDPDSSFMKSI	MTATQLRDLATVYTTTLRYQNPFCEPSRNRTAVSEFMKN	359
Db	301	ARRVSTSTIAYRDPDSSFMKSI	MTATQLRDLATVYTTTLRYQNPFCEPSRNRTAVSEFMKN	360

Qy 360 THVLIRNETPTVYIGTLDMSLLYNETWFVENKTASDNKTTPTSPSGMFORTFIDPLWD 419  
 Db 361 THVLIRNETPTVYIGTLDMSLLYNETWFVENKTASDNKTTPTSPSGMFORTFIDPLWD 420  
 Qy 420 YLDSLLFLDEIRNFSLSRPTVYVNLTPPEHRAVNLSTNSLWMLQ 465  
 Db 421 YLDSLLFLDEIRNFSLSRPTVYVNLTPPEHRAVNLSTNSLWMLQ 466

RESULT 2  
 I38344  
 titin, cardiac muscle [validated] - human  
 N:Alternate names: connectin  
 N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.1-)  
 C:Species: Homo sapiens (man)  
 C>Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004  
 C:Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393  
 R:Labeit, S.; Kolmerer, B.  
 Science 270, 293-296, 1995  
 A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.  
 A:Reference number: A57430; MUID:96026330; PMID:7569978  
 A:Accession: I38344  
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL  
 A:Molecule type: mRNA  
 A:Residues: 1-26926 <LAB1>  
 A:Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:g1017424; PID:g1017425  
 R:Musco, G.; Triatzios, C.; Schuck, P.; Pastore, A.  
 Biochemistry 34, 553-561, 1995  
 A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix m  
 A:Reference number: I38345; MUID:95119041; PMID:7819249  
 A:Accession: I38345  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1977-2014 <MUS>  
 A:Cross-references: EMBL:X83270; NID:9602579; PIDN:CAA58243.1; PID:9602580  
 A:Note: conformation and properties are reported for a synthetic peptide corresponding t  
 R:Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.  
 EMBO J. 11, 1711-1716, 1992  
 A:Title: Towards a molecular understanding of titin.  
 A:Reference number: S20897; MUID:92258380; PMID:1582406  
 A:Accession: S20898  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 13597-14200, 'I', 14202-14696 <LAB2>  
 A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193  
 A:Accession: S20897  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 16330-16382, 'S', 16384-16756, 'P', 16758-16860 <LAB3>  
 A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191  
 A:Accession: S20899  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2  
 A:Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195  
 R:Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.  
 J. Mol. Biol. 256, 556-563, 1996  
 A:Title: Genomic organization of M line titin and its tissue-specific expression in two  
 A:Reference number: S63665; MUID:96177761; PMID:8604138  
 A:Accession: S63665  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 26729-26825 <KOL>  
 A:Cross-references: EMBL:X92412; NID:g1236761  
 R:Gautel, M.; Leonard, K.; Labeit, S.  
 EMBO J. 12, 3827-3834, 1993  
 A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentia  
 A:Reference number: S37393; MUID:94008990; PMID:8404852  
 A:Accession: S37393  
 A:Molecule type: mRNA  
 A:Residues: 26831-26926 <GAU>  
 R:Improta, S.; Polltoun, A.S.; Pastore, A.  
 submitted to the Brookhaven Protein Data Bank, February 1996

A:Reference number: A66736; PDB:1TIT  
 A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341  
 R:Pfuhl, M.; Pastore, A.  
 submitted to the Brookhaven Protein Data Bank, August 1996  
 A:Reference number: A66201; PDB:1NCT  
 A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155  
 C:Genetics:  
 A:Gene: GDB:TTN  
 A:Cross-references: GDB:127867; OMIM:188840  
 A:Map position: 2q31-2q32  
 C:Function:  
 A:Description: structural protein forming filaments in striated muscle  
 C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro  
 C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco  
 structural protein  
 F:24752-25008/Domain: protein kinase homology <KIN>  
 F:84,177,905,2276,2378,2459,2481,2563,2569,2763,2896,3088,3179,3384,3432,3628,3772,4068,  
 98,11066,11486,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1  
 tatus predicted  
 F:16780,16976,17579,17602,17681,17845,17899,18121,18188,18209,18336,18670,18680,18  
 .21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248  
 F:26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental  
 Query Match 27.8%; Score 681; DB 1; Length 26926;  
 Best Local Similarity 1.0%; Pred. No. 0.00036;  
 Matches 268; Conservative 91; Mismatches 106; Indels 25559; Gaps 84;  
 Qy 1 MGRK----- 4  
 Db 649 VGKAEAVATVAANDQAVRREPGRHLESAQAQTTLLEYGYKERISAQVAEPQPPA 708  
 Qy 5 ----- 4  
 Db 709 SEPHVVPKVPRIQAPSETHIKTTQDKGMHSSQIKKTTDLTLRLVHDKRPTASP 768  
 Qy 5 ----- 4  
 Db 769 HFTVSKISVPKTEHGEYASAGSAIATFLQKLSATSAQKITSVKAPTVPKSETRVRAE 828  
 Qy 5 ----- 4  
 Db 829 PTPLPQFPFADPTPYKSEAGVEKVEGVSICTTVREERFVHLHGREAKVTETARVPA 888  
 Qy 5 ----- 4  
 Db 889 PVEIPVTPPTLVSLGNVTVEGESVTLECHISGYPSTVTWYREDYQIESSIDFQITPQ 948  
 Qy 5 ----EMVRD----- 10  
 Db 949 SGIARLMIREAFADSGRFTCSAVNEAGTVSTCYLAVQVSEEPKEKETTAVTEKFTTEEK 1008  
 Qy 11 ----- 10  
 Db 1009 RFVESRDVMTDTSLTEEQAGPGPAAPYFITKPVQKLVEGGSVVFQCGVGNPKPHVY 1068  
 Qy 11 ----- 10  
 Db 1069 WKSQVPLTTGYRYKVSYNKQTGECKLVISMTFADDAAGEYTIIVVRNHGETSASASLEE 1128  
 Qy 11 ----- 10  
 Db 1129 ADYELLMKSOEMLYQVTAFOEPEVGETAPGCVYSEYEKEYEQALIRKMAKQDTV 1188  
 Qy 11 ----- 10  
 Db 1189 VRTVVEDQEPHISFEERLIKIEYRIKTTLELLEDGEKXAVDISEAVESEGFD 1248  
 Qy 11 -----VPK----- 13  
 Db 1249 LRKNRYILEGMVTFHCMSGYPLPKIATWYDKGRINHGERYQMDFLQDGRASLRIPVV 1308  
 Qy 14 ----- 13



```
Db 1309 LPEDEGIYAFASNIKGNACSGKLYVEPAAPLGAPTYIPTLEPVSRIRSLSPRSVRSP 1368
Qy 14 -----MFVLISISFLLVSPINC 30
Db 1369 IRMSPARMSPARMSPARMSPARMSPARMSPARMSPARMSPARMSPARMSPARMSPARMS 1428
Qy 31 -----KUMSK-----35
Db 1429 RFDLVKVGRRPMPETFWFHDGOQIVNDYTHKVVIKEDGTQSLIIVPATPSDSGEWTVVAQN 1488
Qy 36 -----35
Db 1489 RAGRSSISVILTVEAVEHQVPMFVEKLVKNVNIKEGSRLEMKVRATGNPNPDIVWLKNSD 1548
Qy 36 -----35
Db 1549 IIVPHKPKIRIEGTGKEAALKIDSTVSDSAWYTATAINKAGRTDTRCKNVVEVPAEP 1608
Qy 36 -----35
Db 1609 EPERKLIIPRGTYRAKEIAPELEPLHLRYGOEQWEGDLYDKSKQKQPFKXKLTSLRL 1668
Qy 36 -----35
Db 1669 KRFGPAHFECRLTPISDPTMVVWVWLHDGKPLEAANRLRMINEPFGYCSLDYGVAYSRDSGI 1728
Qy 36 -----35
Db 1729 ITCRATNKYGTHTSATLIVKDEKSLVESQLPEGRKGLQRIEELERMAHEGALTGVTTD 1788
Qy 36 -----ALYNRPW-----43
Db 1789 QKEQKQPDIVLYPEVRVLEGETARFCRVGTGYPOPKNWYLNGLIRKSRFRVYDGI 1848
Qy 44 -----GLVLSKI-----50
Db 1849 HYLDIVCKSYDTEGVKVTAEENPEGVIEHKVKLEIQOQREDPRSVLRRAPEPRPFHVHEP 1908
Qy 51 -----50
Db 1909 GKLPQEVQKDRPVDTTTETKEVVKLKRAERITHEKVPESSEBELSKFKRTEEGYEAIT 1968
Qy 51 -----50
Db 1969 AVELSKRKDESYBELLKTKDELHMTKELTEBEKKALABEGKITPTFKPKDKIELSPS 2028
Qy 51 -----50
Db 2029 MEAPKIFERIOSQTVGQSDAHFRVVRVVGKPDPECEWYKNGVKJERSDRIYWYWPEDNVC 2088
Qy 51 -----50
Db 2089 ELVIRDVTAEDSASIMWKAINAGETSHAFLLVQAKQLITFTQELQDVVAKEKDTMATP 2148
Qy 51 -----50
Db 2149 ECETSEPFVKWKYKDGMEVHEGDKYRMHSDRKVHFLSILITDTSADYSCVLVEDENV 2208
Qy 51 -----GKYKL 55
Db 2209 KTTAKLIVEGAVFVKELQDIEVPESYSGELECIVSPENIEGKWHYHNDVELKNGKTYI 2268
Qy 56 -----DOLK 59
Db 2269 TSRGRQNLTKVDVTKEQGEYSFVIDGKTKTKLKMKPRPIALLOGLSDOKVEGDIVQ 2328
Qy 60 LEILRQLET-----TI-----70
Db 2329 LEVKVLSLEVGVMMKQGEVQPSDRVHIVIDKQSHMLLIEDMTKEDAGNYSFTIPALGL 2388
Qy 71 -----70
Db 2389 STSGRVSVSYVDVITPLKDVNVNIEGTKAVLECKVSPDVTSVKRYLNDEQIKPDDRQAI 2448
```

---

```
Qy 71 -----70
Db 2449 VKGTQORLVINRTHASDEGPYKLIIVGRVETNCNLSVEKIKIIRGLRDLTCTETQNVPEV 2508
Qy 71 -----STKYNVS-----77
Db 2509 ELSHSGIDVLNFWKDKKEIKPSSSKYKIBAHGKIYKLVNLMMKDDGKTYTFVAGENMTSGK 2568
Qy 78 -----77
Db 2569 LTVAGGALSPLTDQTVABSQEAVPECEVANPDNPKGEWLDRDGKHLPLTNINRSESDGHR 2628
Qy 78 -----KQPVKNLTM-----86
Db 2629 RLIIAATKLODDIGEYTYKVATSKTSAKLVBAVKIKTKLKNLTVTETQDAVFTVELTHPN 2688
Qy 87 -----86
Db 2689 VKGVOMIKNGVWLESNEKYAISVKGTIYSLRIKCAIVDESUYGFRGLRGASARLHVET 2748
Qy 87 -----86
Db 2749 VKIIKPKDVTALENATVAPEVSVDHTVPVKWPHKSVEIKPSDKGRLVSRKVKHKLMLQ 2808
Qy 87 -----NTEFPQ-----92
Db 2809 NISPSDAGEYTAVVVGQLECKAKLFVETLHTKTKMKNIEVPETKTASFECEVSHFNVPMSW 2868
Qy 93 -----92
Db 2869 LKNGVEIEMSEKPIVVQGLHQLIIMNTSTEDSAEYTFVCGNDQVSATLTVTPIMTSM 2928
Qy 93 -----92
Db 2929 LKDINASEKDTITFEVTVNYEGISYKWLKNGVEIKSTDKCOMRTKLTSLNIRNVHFGD 2988
Qy 93 -----92
Db 2989 AADYTFVAGKATSTATLVVEARHIEFRKHIDIKVLEKKRAMFECEVSEPDITVQWMKDD 3048
Qy 93 -----YILAG-----98
Db 3049 QELQITDRIKIQEKYVHRLIIPSTRMSDAGKYTVWAGNVSTAKLFVEGRDVRIRSICK 3108
Qy 99 -----98
Db 3109 EVQVIEKQRAVVEPEVNEDDVDHWHYKDGIEINFQVQERHKYVVERRIHRMFISETROSD 3168
Qy 99 -----98
Db 3169 AGEYTFVAGNRSSVTLYVNAPPPQVLQELQPVTVQSGKPARFCAMISGRPQPKISWYK 3228
Qy 99 -----98
Db 3229 EEQLLSTGFKCKFLHDQOEYTLIIIEAFPDAAVTCEAKNDYGVATTSASLSVEVPEW 3288
Qy 99 -----PIQN-----102
Db 3289 SPDQEMPVYPPIITPLQDQTVTSEGQPARFCRVSGTDLKVSWYSKOKKIKPSRPFRTMQ 3348
Qy 103 -----102
Db 3349 PEDTYQLEIAEAYPEDESGTYTFVANNVAGQVSSSTANLSLEAPEILHERIEQIEIEMWKE 3408
Qy 103 -----102
Db 3409 FSSSFLSAEBGLHSLQLSKINETLELASESPVYPTKFDSEKGTGPIFIKEVSNADI 3468
Qy 103 -----102
Db 3469 SMGDVATLSVTVIGIPKPKIQWFFNGVLLTSPADYKVFDFGDHSLIILFTKLEDEGEYT 3528
```

Qy 103 ----- 102  
Db 3529 CMASNDYKTTICSAVLKINSKGEGHKDTETESAVAKSLEKLGPCPPHFLKELPIRCAQ 3588  
Qy 103 -----TVLWF----- 110  
Db 3589 GLPAIFEYTVVGEPAFTVTFKENKQLCTSVYTTIIHNPNGSGTFIVNDPOREDSGLYIC 3648  
Qy 111 -----DFVSTOLRKA----- 121  
Db 3649 KAENMLGESTCAAEELLVLEDDTMDTPCKAKSTPEAPEDFPQPLKGPFAVEALDSQOEI 3708  
Qy 122 ----- 121  
Db 3709 ATFFVKDTILKAALITEENQQLSYEHAKANELSSQLPLGAQELQILLEDKLTPESTREF 3768  
Qy 122 ----- 121  
Db 3769 LCINGSIHFOPLKEPSPNLQIIVOSQKTFSGEILMPBEPETOAVLSDDTEKIFPSAMSI 3828  
Qy 122 ----- 121  
Db 3829 EOINSLTVEPLKTLAEPENYPOSSIEPPMHSYLTSVABEVLSEKKTVDTNRQRVT 3888  
Qy 122 ----- 121  
Db 3889 LQKQEQASALILSQSLABGHVESLQSPDWMISQVNYEPLVPSEHSGTEGGKILIESANPL 3948  
Qy 122 ----- 121  
Db 3949 ENAGQSAVRIEEGKSLRFPPLALBEKVLLKEEHSNDVMPDQIIESKREFAIKKVOE 4008  
Qy 122 ----- 121  
Db 4009 VQGRDLLSKESLLSGIPBEOQLNLKIQICRALQAQAVASEQGLFSEWLNRNIEKVEVEAVN 4068  
Qy 122 ----- 121  
Db 4069 ITQEPRHIMCVLYTSAKSVTEEVTTIIEDVDVDPQMANLKMELDALCAIIVEEDILTAE 4128  
Qy 122 ----- 121  
Db 4129 GPRIOQAKTSLQEBMDSFGSQKVEPITEPEVESKYLISTEVSFVNVQSRVKYLDATP 4188  
Qy 122 ----- 121  
Db 4189 VTKGVASAVVDEKQDESLLKPSSEKBSSESGTEEVATVKIQEABGGLIKEDGPMIHTP 4248  
Qy 122 ----- 121  
Db 4249 LVDIVSEBGDIVHLTTSITNAKEVNWYFENKLVPSDEKFKCLQDQNTVTLVIDKVNTEDH 4308  
Qy 122 --KYVSYNNHTAKTI----- 135  
Db 4309 QGEYVCEALNDSGKTATSAKLTWVKRAAPVIRKIEPLEVALGHAKFTCEIQSAPNVRF 4368  
Qy 136 ----- 135  
Db 4369 QMFKAGREIYEDSKSIRSSKYISLEILRTQVVDVCGEYTCASNEYSGVSCTATLTVTV 4428  
Qy 136 ----- 135  
Db 4429 PGGEKVKRKLPERKPEKEEVVLKSVLRKPEBEEPKVEPKLEKVKKPAVPEPPPKP 4488  
Qy 136 -----TFRPPP----- 141  
Db 4489 VEEVEVPTVTKRERKIPBTKVPEIKPAIPLPAPEPKPBEAEVKTIKPPVPEPEPTPIA 4548  
Qy 142 ----- 141  
Db 4549 APVTVVVGKKAENAKAPKEEAAKPGPIKGVPKTTPSIEARRKLRPGSGGKPPDEAP 4608  
Qy 142 ----- 141

Db 4609 FTYQLKAVPLFKVEIKDIIILTESEFVGSSAIFCELVSPSTAITWTMKGDSNIRESPKHR 4668  
Qy 142 ----- 141  
Db 4669 FIADGDKRLHIIDVQLSDAGEYTCVLRGNKERTSTAKLVEELPVRVFKTLSEEVTVV 4728  
Qy 142 -----C-----GRV----- 145  
Db 4729 KGQPLYLSCELNKERDVVWRKDGKIIVVEKPGRIVPGVIGLMRALTINDADDDTAGTYTVT 4788  
Qy 146 ----- 145  
Db 4789 VENANNECSCVKVVEVIRDWLKPIRDQHVKPGTAIFACDIAKTPNIKWFKGYDEI 4848  
Qy 146 ----- 145  
Db 4849 PAEPNDKTEILRDGNHLYLKIKNAMPEDIAAYAVEIEGKRYPAKLTJGEREVELLKPIED 4908  
Qy 146 -----PSMTC----- 150  
Db 4909 VTIYKESASFDAIEISADIPGQWKLKCELLRPSPTCEIKAEGBGRFLTLLHKVKLDQAGE 4968  
Qy 151 ----- 150  
Db 4969 VLYQALNAITTAILTAVKEIELDFAVPLKDVTVPERQARPECVLTREANVWSKGPDIK 5028  
Qy 151 ----- 150  
Db 5029 SSDKFDIIADGKXHLVINDSQFDDGYYTAEVEGKTSARLFTVGTIRLKPMSPLEDQTV 5088  
Qy 151 ----- 150  
Db 5089 KEGETATFVCELSHEKMHVWFKNDAKLHTSRTLISSEBCKTHKLEKVEVTLDDISQIKA 5148  
Qy 151 ----- 150  
Db 5149 QVKELSTAQLKLEADPYFTVKLHDKTAVEKDEITLKCEVSKDVPVKWFKDGEIIVPSP 5208  
Qy 151 ----- 150  
Db 5209 KYSIKADGLRRIILKIKKADLKDKGEYVDCGDTKTKANVTVEARLIEVEKPLYGVEVPVG 5268  
Qy 151 ----- 150  
Db 5269 ETAHFIEELSEPDVHGQWKLKGQPLTASPDCEIIEDGKXHLILHNCQLGWTGEVSFQAA 5328  
Qy 151 ----- 150  
Db 5329 NAKSAANLKVKEPLFIITPLSDVKVFEKDEAKPECEVSREPKTFRWLKGTQBITGDORF 5388  
Qy 151 ----- 150  
Db 5389 ELIKDGTKHSNVTKSAFAFEDEAKYMFABDKHTSGKLIIEGIRLKFLLTPLKDVTAKEKES 5448  
Qy 151 ----- 150  
Db 5449 AVFTVELSHDNIRVKWFKNDORLHTTRSVSQDEGKTHSITFKDLSIDDTSQIRVEANGM 5508  
Qy 151 ----- 150  
Db 5509 SSEAKLTVLEGDPYFTGKLQDYTGVEKDEVILQCEISKADAPVKWFKDGEIKPKSKNAV 5568  
Qy 151 ----- 150  
Db 5569 KTDGKKRMLILKXALKSDIGQYTCDCGTDKTSKGLDIEDREIKLVRPLHSVEMETETAR 5628  
Qy 151 ----- 150  
Db 5629 PETEISEDDIHANWKLKGEALLQTPDCEIKEEGKIHSLVHLNCELDDQGTGGVDFOAANVKS 5688  
Qy 151 ----- 150

Db 5689 SAHLRVKPRVIGLLRPLKDVTTAGETATFDCELSYEDIPIVWYLGKKLEPDKVPR 5748  
QY 151 ----- 150  
Db 5749 EGKVHTLRLDVKLEBAGEVOLTAQDKFKTHANLPVKEPPVBFTKPLEDQTVVEGATAVLE 5808  
QY 151 ----- 150  
Db 5809 CEVSRENAKVWFKNGTILSKKYEIVADGRVKLVIHDCPTEDIKTYTCDAKDKFTSC 5868  
QY 151 ----- 150  
Db 5869 NLNVPPHVEPLRLDLOVREKENARFECELSRENAKVKWFKDGAIEKKKKYDIISKG 5928  
QY 151 ----- 150  
Db 5929 AVRILVINKLLDDEABEYCEVRTARTSGMLTVLEBEAVFTKNLANIEVSETDTIKLVCE 5988  
QY 151 ----- 150  
Db 5989 VSKGAEVIWYGDDEIETCRYEILTEGRKRILVIONAHLEDAGNYNCRLPSSRTDGK 6048  
QY 151 ----- 150  
Db 6049 KVHELAAEFISKQWLEILEGEKABFVCSISKESPPVQWKDDDKTLESGDKYDVADGKK 6108  
QY 151 ----- 150  
Db 6109 RVLVVKDATLQDMGTVMVVGAAAHILTIEKLRIVVPLKDRVKEQEVFNCEVNT 6168  
QY 151 ----- 150  
Db 6169 EGAKAKWRNBEAIPDSSKYIILQKDLVYTLIRDAHLDDQANVNSLTHRGENVKSAA 6228  
QY 151 ----- 150  
Db 6229 NLIVEEDLRIVEPLKDIETMEKKSVTFWCKVNRNLNVLKWTXNGEEVFFDNRVSYRVDK 6288  
QY 151 ----- 150  
Db 6289 YKHLMTIKDCGPDGSEYIVTAGQDKSVAELLIIIEAPTFFVEHLEDQTVTFDDAVFSCQ 6348  
QY 151 ----- 150  
Db 6349 LSREKANVWYRNGREIKEGKKYFPEKDGSIHRLIIKDCRLDDECEYACGVEDRKSARL 6408  
QY 151 ----- 154  
Db 6409 FVEEIPVEIIRPPQDILEAPGADVFLAELNKKVQVQWLNRRNMVVGDKHQMSEGI 6468  
QY 155 ----- 154  
Db 6469 HRLQICDIKPRDQGEYRFIADKEARAKLELAAPKIKTADQDLVVDVGKPLTWVVPYDA 6528  
QY 155 ----- 154  
Db 6529 YPKAEAWFKENEPLSTKTIIDTAEQTSFRILEAKKGDKRYKIVLQNHKGABGFNLK 6588  
QY 155 ----- 154  
Db 6589 VIDVPGVRNLEVTETFDGEVSLAWEBPLTDGGSKIIGYVVERRDIKRKTWVLTADRAES 6648  
QY 155 ----- 157  
Db 6649 CEFTVTGLQGGVEYLFVRSARNVGTGEPVETDNPEARSKYDVPGLPLNVTTIDVNR 6708  
QY 158 ----- 157  
Db 6709 GVSUTWEPPEYDGAIEINYYIELRDKTSIRWDTAMTVRAEDLSATVTVVVEGQEYSFRV 6768  
QY 158 ----- 157  
Db 6769 RAQNRIGVKPFAATPFVKVADPIERPSPVNLTSDDQTSVQLKWEPLPKDGGSPILG 6828

QY 158 ----- 157  
Db 6829 YIIBCEBGDNWIRCNMKNLVPCLTYKVTGLEKGNKYLRYSAENKAGVSDPSEILGLPLT 6888  
QY 158 ----- 157  
Db 6889 ADDAPVEPTMDSAFKOGLEVI VNPITILVPSTGYPRPTATWCFGDKVLETGDRVKWMT 6948  
QY 158 ----- 157  
Db 6949 LSAYAEVLVSPERSDKGIYTLKLENRVKTTISGEIDVNVNVIARPSAPKELKFGDITKDSVH 7008  
QY 158 ----- 157  
Db 7009 LTWEPDDGGSPLTGYVVEKREVSRTWTWKVMDFTDLBFTVPLVQCKEYLFKVCARN 7068  
QY 158 ----- 157  
Db 7069 KCGEPAYVDEPVNMSTPATVPDPENNVKWRDRTANSIFLTWPPKNDGSRKGYI V 7128  
QY 158 ----- 157  
Db 7129 RCPRGSKWVACGEPVAETKMEVTGLBEGKYYAYRVKTLNRQGASKPSRPTTEEIOAVDTQ 7188  
QY 158 ----- 157  
Db 7189 EAPBIFLDVLLAGLTVKAGTKIELPATVTKGPEPKITWTKADMILKQDKRITTIENVPK 7248  
QY 158 ----- 164  
Db 7249 STVTIVDSKRSDDTGTYYIEAVNVCGRATAVVEVNVLDKPGPPAAPDITDVTNESCLL 7308  
QY 165 ----- 164  
Db 7309 PPRDDGSKITNYVVERRATDSEVWHKLSSTVKOTNPKATKLI PNKEYI PRVAENMYGA 7368  
QY 165 ----- 164  
Db 7369 GEPVQASPI TAKYQDPGPPPTLRLEPSDITKDAVTLTWCEPDDGGSPITGVVVERLDP 7428  
QY 165 ----- 164  
Db 7429 TDKVRCNMPVKDITYRVKGLTNKKYRFRVLAENLAGPGKPSKSTEPILIKPIDPPW 7488  
QY 165 ----- 164  
Db 7489 PPGKPTVKDVGKTSVRLNWTXPEHDGGALESYVVIEMLKGTGDEWVRVAEGVPTQHLLP 7548  
QY 165 ----- 164  
Db 7549 GLMEGQEYSFRVAVNKAGESESPSPDVLCREKLYPPSPRWLEVINIKNTADLKW 7608  
QY 165 ----- 164  
Db 7609 VPEKGGSPITNYIIVEKRDVRRKGWQTVDTTVKTKCTVTPLTEGSLYVFRVAENAIGQ 7668  
QY 165 ----- 164  
Db 7669 SDYTEIEDSVLAKDTFTTPGPYPYALAVVDVTKRHVDLKWEPKNDGGRPIORYVIEKKER 7728  
QY 165 ----- 169  
Db 7729 LGTRWVKAGTAGPDCNFRVTDVIEGTEVQFQVRAENEAGVGHFSEBTEILLSIEDPTSP 7788  
QY 170 ----- 169  
Db 7789 SPPLDLHVTDAGRKHIAIAWKPEKNGSGPIIGYHVEMCPVGTCKWVRNRPINKDLKFK 7848  
QY 170 ----- 169  
Db 7849 VEEGVVPDKYVLRVRAVNAIGVSEPSSEISENVVAKDPDCKPTIDLETHDIIVIEGEKLS 7908

QY 170 ----- 169  
Db 7909 IPVPFRAVPVTVSMHKDGEKVASDRLTMKNDHISAHLEVPKSVRADAGIYITILENKL 7968  
QY 170 ----- 169  
Db 7969 GSATASINVKVIGLPGCKDIKASDITKSSCKLTWEPPDFDGGTPIHLHYLERREAGRR 8028  
QY 170 ----- 169  
Db 8029 YIPVMGENKLSWTVKDLIPNGEYFVRKAVNKVGGGYIELKNPVIADQPKQPPDPVD 8088  
QY 170 ----- 169  
Db 8089 VEVHNPATAEMTITWKPPLYDGGSKIMGYIIETIAKGEHRWKRCNEHLVPILTYYTAKGLE 8148  
QY 170 ----- 169  
Db 8149 EGKEYQFRVRAENAAGISEPSRATPPTKAVDPIDAPKVILRTSLEVKGDEIADASISG 8208  
QY 170 ----- 169  
Db 8209 SPYPTITWIKDENVIPBEIKKRAAPLVRRRKGEVQEEEPVLPLOTQLSIDNSKKGESQ 8268  
QY 170 ----- 169  
Db 8269 LRVRDSLRLPDHGLYMIKVENDHGIAKAPCTVSVLDTGPPINFVFPEDIRKTSVLCKWEPP 8328  
QY 170 ----- 172  
Db 8329 LDDGSEIINTLEKKDKTKPDSEWIVVTSLRHCKYSVTKLIEGKEYLFRVRAENRFGP 8388  
QY 173 -----TFNPMF----- 178  
Db 8389 GPPCVSKPLVAKDPGPPDAPDKPIVEDVTSNMLVKWNEPKDNGSPILGWLEKREVNS 8448  
QY 179 ----- 178  
Db 8449 THWSRVNKSLLNALKANVDGLLEGLTYVFRVCAENAGPGKFSPPSDPKTAHDPISPGPG 8508  
QY 179 ----- 178  
Db 8509 PIPRVTDTSSTIELEWEPAPNGGGEIVGYVDKQLVGTNKSRCRTEKMIKVRQYTYKE 8568  
QY 179 ----- 178  
Db 8569 IREGADYKLRVSANNAAGEGPPGETQPVTVABPQEPAPVELDVSVKGIQIMAGKTLRIP 8628  
QY 179 ----- 178  
Db 8629 AVVTGRPVPTKVWTKGEGELDKRVVIDNVGTKSELIIKDALKDGHGRYVITATNSCGSK 8688  
QY 179 ----- 178  
Db 8689 FAAARVEFDVPGPVLCLKPVVTRNRMCLNWSDEDDGSEITCFIIERKDKAMHTWRQ 8748  
QY 179 ----- 178  
Db 8749 PIETERSKCDITLLEGOEYKFRVIKAKNKGCGPPVEIGPILAVDPLGPPTSPELTYTE 8808  
QY 179 ----- 178  
Db 8809 RQRSTITLDWKEPRNGSGSPIQGYIIKKRRHDKPDERVNRKLCPTTSTFLVENLDEHQMY 8868  
QY 179 ----- 178  
Db 8869 EFRKAVNEIGESBPSLPLANVVIQDDEVPPTIKLRLSVRGTTIKVKAGEPVHIPADVTGL 8928  
QY 179 ----- 178  
Db 8929 PMPKLEWSKNETVIEKPTDALQITKEEVSRSSEAKTELSIPKAVREDKGTYYTIVTASNRLGS 8988  
QY 179 ----- 178

Db 8989 VFRNVHVEVDRPSPRPNLAVTDIKABSCYLTWAPLDNGSGSEITHYVIDDKRDASRKA 9048  
QY 179 ----- 178  
Db 9049 WEEVTNTAVEKRYGIWKILPNGOYEFVRVAVNKYGISDECKSDKVVIQDPYRLPGPPGPK 9108  
QY 179 -----PNV 181  
Db 9109 KVLARTKSMVLVSWTPPLDNGGSPITGWLEKREBEGPYWSRVSRAPIITKVGLKGVEFNV 9168  
QY 182 PR----- 183  
Db 9169 PRLLGKVYQFRAMAINAAGIGPPSESDPEVAGDPIFPFGPPSPCEVKDKTKSSISLGW 9228  
QY 184 ----- 183  
Db 9229 KPPAKDGGSPIKGYIVEMQEEGTTDWKRVNBPDKLITTCVCVVPNLKELRYFRVKA 9288  
QY 184 ----- 183  
Db 9289 EAGESESDTTGEBIPATDIOEEPEVFDIAGQDCLVCKAGSQIRIPAVIKGRPTPKSSWE 9348  
QY 184 ----- 183  
Db 9349 FDGKAKKAMKDVHDIPEDAQLETAENSSVIIPECKRSHTKYSITAKNKGQKTANC 9408  
QY 184 ----- 183  
Db 9409 VKVMDVPGPKDLKVSIDITRGSCRLSWKMPDDDDGDRIKGYVIEKRTIDGKAWTKVNPDC 9468  
QY 184 ----- 183  
Db 9469 GSTTFVVVDLLSEQOYFVRVRAENRFGIGPPVETIQRRTARDPIYPPDPPIKLKIGLITK 9528  
QY 184 ----- 183  
Db 9529 NTVHLSWKPKNDGSGPVTHYVIECLAWDPTGTKEARQCNKRDVELOFTVEDLVEGG 9588  
QY 184 ----- 183  
Db 9589 EYEFVRKAVNAAGSVKSPSATVGPCDQCPDMPSPSIDLKBFMEVEBEGTNVIVAKIKGVPF 9648  
QY 184 ----- 183  
Db 9649 PTLTWFKAPKKCPDNKEPVLVDTHVNLVVDCTCLVIPOSRRSDTGLYTTITAVNNLGTA 9708  
QY 184 ----- 183  
Db 9709 SKENRLNVLGRPGPPGVPIKFSVSADQMTLSWPPPKDGGSKITNVYIEKREANRKTW 9768  
QY 184 ----- 183  
Db 9769 HVSSEPKECTYTIPLKLEHGYVFRIMAQNKYIGEGEPLDSEPETARNLFSVPGAPDKPTV 9828  
QY 184 ----- 183  
Db 9829 SSVTRNSMTVNWEEPYDGGSPVTGYWLEKMDTTSKRWRKVRNRPDIKAMTLGVSKVTGL 9888  
QY 184 ----- 183  
Db 9889 IEGSDYQFRVVAINAAAGVPSLPDPATARDPIAPGPPPKVTDWTWKSSADLEWSPL 9948  
QY 184 ----- 183  
Db 9949 KDGSKVTGYVIEYKEBEGKEWKGDKVRGTVLVTGLKBGAFYKFRVSAVNIAGIGE 10008  
QY 184 ----- 183  
Db 10009 PGEVTDVIEMKDRIVLQDASVRDRIVVHAGGVIIRIIAYVSGKPPPTVTWNMERTL 10068  
QY 184 ----- 183

```
Db 10069 PQEATIEITAISSSMWIKNCQSHQGVYSLAKNEAGERKKTIIVDVLDPGPVGTPELA 10128
Qy 184 ----- 183
Db 10129 HNLTNESCKLTWSPEDDGGSPITWYIEKRESDRRAWTPVYTVTRQNAVQGLIQGA 10188
Qy 184 -----WNTKLYG- 191
Db 10189 YFFRIAENSIGMGPFVETSEALVIREPITVPERPEDLEVKEVTQNTVTLTNPPKYDGG 10248
Qy 192 ----- 191
Db 10249 SEIINYVLESRLIGTEKFKHTNDNLLSRKTVTKGDKGDTYEVRSAVNIVGQKPSFC 10308
Qy 192 ----- 191
Db 10309 TKPITCKDELAPPTLHLDPRDKLTIRVGEAFALTGRYSGKPKVSWFKDEADVLEDDRT 10368
Qy 192 ----- 191
Db 10369 HIKTPATLALEKIKAKRSDSGKCYVVVENSTGSRKGFQCVNVVDHPGPPVGPVSFDEVT 10428
Qy 192 ----- 191
Db 10429 KOYMWISWKPPPLDDGSKIYIYIEKKEVGKDVMPVTSASAKTCKVSKLLEGGDYIFR 10488
Qy 192 ----- 191
Db 10489 IHAENLYGSDPLVSDSMKAKDRPRVPDAPDQPIVTEVKDSALVTWKNKPHDGGKPTNY 10548
Qy 192 ----- 191
Db 10549 ILEKRETMKRWARTKDIHPYTKFRVPDLLEGQCYEFRVSAENEIGIGDPSPPKPVF 10608
Qy 192 ----- 191
Db 10609 AKDPIAKSPPPNPEAIDTTCNVDLTWQPPRHGGSKILGYIYEYQKVGDEEWRANHT 10668
Qy 192 ----- 191
Db 10669 PESCPETKYVTGLRDGQTKYFRVLAVNAAGESDPAHVPEPVLVKORLEPPELILDANMA 10728
Qy 192 -----PTKVNVD- 198
Db 10729 REQHIKVGDTLRLSAILKGVPPKVTWKEDRDAPTKARIDVTPVGSKLEIRNAAHEDGG 10788
Qy 199 ----- 198
Db 10789 IYSLTVENPAGSKTVSVKVLVDKPGPPRDLEVSIRKDSCVLTWKEPLDDGGSVITNYV 10848
Qy 199 ----- 198
Db 10849 VERRDASAQWSPLSATSKKSHFAKHLNEGNOYLFRVAENQYGRGPFVETPKPIKALD 10908
Qy 199 ----- 198
Db 10909 PLHPPGPKDLHHVDVDTEVSLVWKNKDRDGGSPITGYLVEYQEBGTDWIKFKVTNL 10968
Qy 199 -----SQIYIFGL- 207
Db 10969 ECVVTGLOQKTYRFRVKAENIVGLPDTTPIECQKLPVPSVELDVKLIEGLWVKAG 11028
Qy 208 ----- 207
Db 11029 TTVRPAILRGVPVPTAKWTTDGSSEIKTDEHYTVETDNFSSVLITIKNCLRDRDTGEYQITV 11088
Qy 208 ----- 207
Db 11089 SNAAGSKTVAVHLTVLDVGPPTGPINILDVTPHEMTISWQPPKDDGGSPVINIVEKQD 11148
Qy 208 -----TALLRYAQBN- 219
Db 11149 TRKDTGVVSSGSSKTKLKIPLHQGCBYVFRVRAENKIGVGPPLDSTPTVAKHKFSPPS 11208
```

---

```
Qy 220 ----- 219
Db 11209 PPGKPVVTDITENAATVSWTLPKSDGGSPITGYTMERREVTGKWVRVNTKPIADLKFRVT 11268
Qy 220 ----- 219
Db 11269 GLYEGNTYFRVFAENLAGLSKPSFSDPIKACRPIKPPGPPINPKLKDKSRETADLWVT 11328
Qy 220 ----- 219
Db 11329 KPLSDGGSPILGYVVEQKPGTAQWNRINKDELIRQCAFRVPGGLIEGNEYFRIKAANIV 11388
Qy 220 ----- 219
Db 11389 GEGSPRELAESVIAKIDILHPPEVELDVTCRDVITVRVQQTIRILARVKGRPEPDITWKE 11448
Qy 220 ----- 219
Db 11449 GKVLVREKRVDLIODLPRVELQIKEAVRADHGKYYII SAKNSSGHAQGSAINVVLDRPGPC 11508
Qy 220 ----- 219
Db 11509 QNLKVTNVTKENCCTIISWENPLDNGSSEITNFI VEYRKPNQKWSIVASDVTKRLIKANLL 11568
Qy 220 ----- 219
Db 11569 ANNEYFRVCAENKVGVGPTIETKTPILAINPIDRPGEPENLHIADKGKTFVYLKWRPDP 11628
Qy 220 ----- 219
Db 11629 YDGGSPNLSYHVERRLKSGDDWERVHKSGIKETHYMDRCVENQIYEFRTQKNEGGESD 11688
Qy 220 ----- 219
Db 11689 WVKTEVVVKEDLQKPVLDLKLSGVLTVKAGDTIRLEAGVKGKPPFVAVTKDKDATDLT 11748
Qy 220 ----- 219
Db 11749 RSPRVKIDTRADSSKFSLTRAKRSDGGKGYVVTATNTAGSFVAYATVNVLDKPGPVRLKI 11808
Qy 220 ----- 219
Db 11809 VDVSSDRCTVCDPPEDDGGCEIQNYILEKETKMWSTYSATVLTFGTTVTRLIEGNE 11868
Qy 220 ----- 219
Db 11869 YIFRVAENKIGTGPPTESKPVIAKTKYDKPGRDPPEVTVKVSKEMTVVWNPPEYDGGK 11928
Qy 220 ----- 219
Db 11929 SITGYFLEKKEKHSTRWVPVNVKSAIPERRMKVQNLDPDHEYQFRVKAENEIGIGEPSLPS 11988
Qy 220 ----- 219
Db 11989 RPVVAKDPIEBPPGPTNFRVVDTTKHSITLGMGKPVYDGGAPIIGYVVMEMPKIADASPD 12048
Qy 220 ----- 219
Db 12049 EGWKRCAAAQLVRKEFTVTSLDENQBYEYFRVCAQNVQIGORPAELKEAIKPKELPEPPE 12108
Qy 220 ----- 219
Db 12109 IDLDASMRKLVIVRAGCPIRLFAIVRGRPAPKVTWRKVIDNVVRKQGVLDVDTMAFLVI 12168
Qy 220 ----- 219
Db 12169 PNSTRDPSGKYSLTFLVNPAGEKAVFVNVRLDTPGPVSDLKVSVDTKTSCHVSWAPPEND 12228
Qy 220 -----TH- -SFLVN- 227
Db 12229 GGSQVTHYVEKREADRKTSTVTPEVKKTSFHVNLNLPVGNYYFRVAVNVEYGGVPTD 12288
```

QY	228	-----	227
Db	12289	VPKPVLASDPLSEPPPRKLEATEMTKNSATLAWLPPLRDGAGKIDGYIISYREBEQPAD	12348
QY	228	-----	227
Db	12349	RWTEYSVVKDSLAVVTGLKEGKKYKFRVAARNAVGLPREAEGVYEAKQLPPKILMP	12408
QY	228	-----	227
Db	12409	EQITIKAGKKLRIEAAHVYKPHPTCKWKKGDEVVTSHLAVHKADSSILIIKDVTRKD	12468
QY	228	-----	227
Db	12469	SGVYSIATENSSGTDQKIKVVVMDAPGPPQPFDISDIDADACSLSMWHIPLDGGSNIT	12528
QY	228	-----	227
Db	12529	NYIVEKDVSRGDWTALASVTKTSRCVGKLIPOQEYIFRVAENRFGISEPLTSPKMA	12588
QY	228	-----	227
Db	12589	QPPFGVPSEPKNARVTKVKNKOCIFVANDRPDSGSGPIIGYLIERKERNLLWKANDTL	12648
QY	228	-----	227
Db	12649	VRSTEYPCAGLVEGLEYSFRIYALNAGSSPPSKTEYVYTABMPVDPGPKPEVIDTKST	12708
QY	228	-----	227
Db	12709	VSLIWARPKHDGSKIIGYVEACKLPGDKWVRCNTAPHQIQBEYATATGLEKAQOFR	12768
QY	228	-----	227
Db	12769	AIARTAVNISPPSESDPVTIIAENVPPRIDLSVAMKSLITVYAGTNVCLDATVFGKMP	12828
QY	228	-----	233
Db	12829	TVSWKKGDTLLKPAEGIRKWAQMRNLCTLELFSVNRKSDGYTITAEENSSGSKSATIKLV	12888
QY	234	-----	233
Db	12889	LDKGPASVKINKMYSDRAMLSWEPLEDGSEITNYIVDKRETSRPNWAQVSATVIT	12948
QY	234	-----	233
Db	12949	SCSVEKLIEGHEYQFRI CAENKYGVDPVFTEPAIAKNPYDPPGRCDPPIVSNITKDHT	13008
QY	234	-----	233
Db	13009	VSWKPPADGGSPITGYLLEKRETQAVNWKVNRKPIIERTLKATGLQEGTEYEFRTAI	13068
QY	234	-----	233
Db	13069	NKAGPKPSDASAAAYARDPQVPAPPAPPKYDYDTRSSVSLSWGKPAYDGGSPIIGYLV	13128
QY	234	-----	233
Db	13129	EVKRADSNVRCNLPQNLQKTRFEVTGLMEDTQYQFRVYAVNKIGYSDPSDVPDKHPK	13188
QY	234	-----	233
Db	13189	DILIPPEGEHADLRKTLILRAGVTMLRYVPVKGRPPPKITWSKPNVNLDRDIGDIKST	13248
QY	234	-----	233
Db	13249	DFDTFLRCENKNYDAGKIITLLENSCGKEYTIVVKVLDTPGPPINVTVKEISKDSAYV	13308
QY	234	-----	233
Db	13309	TWEPPIIDGGSPINIYVVKRDAERKSWSTVTTECKSTSFVRPNLEBCKSYFFRPAENE	13368
QY	234	-----	233

Db	13369	YGIGDPGETRDAVKASQTPGPPVDLKVRSVSKSSCSIGWKPHSGSGSRIIGYVVDFLTE	13428
QY	234	-----	233
Db	13429	ENKQVRVMSLSLOYSAKDLTEGKEYTFRVSAENENGEGTSEITVVARDVDWAPDLDK	13488
QY	234	-----	233
Db	13489	GLPDLCYLAKENSFRLKIPKIGKAPSVSWKKGEDPLATDTRVSVSSAVNTTLIVYDC	13548
QY	234	-----	233
Db	13549	QKSDAGKYTITLKNVAGTKEGTISIKVVGKPGIPTGPIKPDDEVTAEMTLKWAAPPKODGG	13608
QY	234	-----	233
Db	13609	SEITNYILEKRDSVNNKVVTCASAVQKTFTRVTRLHEGMEYTFRVSANENKYGVGGLKSE	13668
QY	234	-----	233
Db	13669	PIVARHPFDVPDAPPPPNIVDVHRDSVSLTWTDPKTKGGSPITGYHLEFKERNLLWKRA	13728
QY	234	-----	236
Db	13729	NKTFIRMRDFKVTGLTEGLEVEFRVMAINLAGVGKPSLPSEPVALDPIDPPGKPEVINI	13788
QY	237	-----	245
Db	13789	TRNSVTLIWTEPKYDGGHKLTYIVEKRDLPFSKWMKANHVNPPECAFTVTDLVEGGKYE	13848
QY	246	-----	245
Db	13849	FRIRAKNTAGAI SAPSESTETIICKDEYEAPTIVLDPTIKDGLTIKAGDTIVLNAISILG	13908
QY	246	-----	245
Db	13909	KPLPKSSWSKAGKDIRPSDITQITSTPTSSMLTIKYATRKDAGEYITITATNPFCTKVEHV	13968
QY	246	-----	245
Db	13969	KVTVDVPGPPGPPVEISNVAEKATLTWTPPLEDGGSPIKSYILEKRETSRLLTWVSED	14028
QY	246	-----	246
Db	14029	IQSCRHVATKLIQONEYIFRVSANVHYKGEVPQSEPVKMVDRPGPPGPEKPEVSNVTK	14088
QY	247	NT-----	253
Db	14089	NTATVSKRPVDDGGSEITGYHVHREKKSILRWVRAIKTPVSDLRCKVTGLQEGSTYEFR	14148
QY	254	-----	253
Db	14149	VSAENRAGIPPPSEASDVLMOAAAYPPGPPSNPHVTDTTKKSASLAWGKPHYDGGLEIT	14208
QY	254	-----	253
Db	14209	GYVVEHQVGEAWIKDTCGTALRITQFVVVDPDLQTKYKNFRISAINDAGVGEPAVIPDV	14268
QY	254	-----	258
Db	14269	EIVEREMAPDFELDAELRRLTVVRAGLSIRIFVPIKGRPAPEVTWTNDINLKNRANIEN	14328
QY	259	-----	258
Db	14329	TESFTLLIIPECNRYDTGKFVMTIENDPAGKSGFVNVVRLDTARPPQLRPTDITKDSVT	14388
QY	259	-----	258
Db	14389	LHWDPLIDGGSRIYNIYVEKREATRKSYSTATTCKHCTYKVTGLSEGCEYFFRVAEN	14448
QY	259	-----	258

Db 14449 EYGIGETETTPVKASEAPSPDNLNMDITKSTVSLAWPKPKUDGGSKITGVIEAQR 14508  
QY 259 -----VKE----- 261  
| | |  
Db 14509 KGSQWTHITTVKGLECVVRNLTEGEYVTFQVMVNSAGRSAPRSPVIVKEQTMPLPEL 14568  
QY 262 ----- 261  
Db 14569 DLRGYQKLVIAKAGDNIKVEIPVLGRPKPTVTKGDOILKQTORVNFETTATSTILNI 14628  
QY 262 ----- 261  
Db 14629 NECVRSDSGPPLTARNIVGEVGDVITIQVHDIPOGPTGPDKFDEVSDFVTFPSWDPPEN 14688  
QY 262 -----QFEKA----- 267  
| | |  
Db 14689 DGGVPISNVVEMRQDSTTWVETLATTVIRTYKATRLTTGLEYOFRVKAQNRVGVGPI 14748  
QY 268 ----- 267  
Db 14749 TSAMIVANYPKVPGPGTPOVAVTKDSMTISMHEPLSDGSPILGYHVERKERNGILW 14808  
QY 268 ----- 267  
Db 14809 QTVSKALVPGNIFKSSGLTDGIAEYFRVIAENMACKSKPSKSPBMLALDIDPPGKVPV 14868  
QY 268 ----- 267  
Db 14869 LNIIRHTVTWKWPEYTGPKITSYIVEKRDLNPNRWLKNFNSILENEFTVSLGTEDA 14928  
QY 268 ----- 267  
Db 14929 AYEPRVIAKNAAGAISSPSPSDAITCRDDVEAPKIKVDVKFOTVILKAGEAFRLEADV 14988  
QY 268 ----- 267  
Db 14989 SGRPPPTMEWSKDGKELEGTAKEIKIADFTNLVNKOSTRRDGSAYTLTATNPGGFANH 15048  
QY 268 ----- 267  
Db 15049 IPNVKVLDRPGPEGLAVTEVTSEKCVLSWPPPLDDGAKIDHVIQKRETSRLAWTNV 15108  
QY 268 ----- 267  
Db 15109 ASEVQVTKLVTKLLKGNIEYIFRVMANVKYGVGEPLSEPLAVNPYPPDPKPNPEVTT 15168  
QY 268 -----KKT----- 270  
| | |  
Db 15169 ITKDSMVVWCHPDSDGSGSEIINYIVERRDDKAGORWIKCNKKTLDLRYKVSGLTEGHEY 15228  
QY 271 -----QSTTTPYF----- 278  
| | |  
Db 15229 EFRIMAENAGISAPSPTSFPYKACDTPKPGPGNPRVLDTSRSSISIANWKPIYDGS 15288  
QY 279 ----- 278  
Db 15289 EITGMVEIALPEDEWQIVPPAGLKATSYITITGLTENQEKIRIYAMNSEGLGEPALV 15348  
QY 279 ----- 278  
Db 15349 FGTPKAEDRMPLPEIELDADLRKVVTIRACCTLRLFPVPIKGRPDPEVKWARDHGESLDKA 15408  
QY 279 -----SVT----- 281  
| | |  
Db 15409 SIESASSYTLIIVGNVNRFDGSKYILTIVENSNGSKSAFVNVRLDTPGPPQDLKVEYTK 15468  
QY 282 ----- 281  
Db 15469 TSVTLTWDPPLLDGSKIKNIYIVEKRESTRKAYSTVATNCHTSKWVDQLQEGCYSYFRV 15528  
QY 282 ----- 281  
Db 15529 LAENEYIGLPAETAESVKASERPLPPGKITLMDVTRNSVSLSWKPEHGGSRILGYIV 15588

QY 282 -----TSAALNVT----- 289  
| | |  
Db 15589 EMQTKGSDKWATCATVKVTEATITGLIQGBEYSFRVSAQNBKIGSDPRQLSVPIAKDLV 15648  
QY 290 ----- 289  
Db 15649 IPPAPKLLFNFTFTVLAGEDLKVDPFFIGRPTPAVTMHWKDNVPLKQTTVRNAESTENSL 15708  
QY 290 ----- 289  
Db 15709 TIKDACREDVGHVVVVKLTNSAGEALETNLVILDKPGPPTGPKVMDDEVTAADSIITLSWGP 15768  
QY 290 ----- 289  
Db 15769 KYDGGSSINNYIVEKRDSTTTTWOIVSATVARTTIKACRLKTGCEYQPRIAENRYGKST 15828  
QY 290 ----- 289  
Db 15829 YLNSEPTVAQYPPKVPGPPTVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERN 15888  
QY 290 ----- 289  
Db 15889 LNVKLNKTPIPQTKFKTTGLEEGVEYEFVSAENIVGIGKPSKYSECYVARDPCDPPGRP 15948  
QY 290 ----- 289  
Db 15949 EAIIVTRNSVTLQWKPTYDGGSKITGYIVEKBELPEGRMKASFNTIIDTHFEVTLVE 16008  
QY 290 ----- 289  
Db 16009 DHRYEPRVIAARNAAGVSESPSESTGAITARDEVDPPIRSMDPKYKDTIVVHAGESFKYDA 16068  
QY 290 ----- 289  
Db 16069 DIYKPIPTQWIKGDOELSNATBLEIKSTDFAISLVKDAVRVDSNGVYILKAKVAGER 16128  
QY 290 ----- 289  
Db 16129 SVTVNVKVLDRPGPGPVISGVTAEBCTLAWKPLQDGGDIINYIVERRETSRLVMT 16188  
QY 290 ----- 289  
Db 16189 VVDANVQTLCKVTKLLEGNEYTPRIMAVNKYGVGEPLSEPLSEPVVAKNPFVVPDAPKAEV 16248  
QY 290 ----- 289  
Db 16249 TTVTKDSMIVVWERPASDGGSEILGYVLEKRDKEGIRTRCHKRLIGELRLRVTGLIENH 16308  
QY 290 ----- 289  
Db 16309 DYBPRVSAENAGLSEPPSPSAYQACDPIYKPGPPNPNKVIDITRSSVFLSKPIYDG 16368  
QY 290 ----- 289  
Db 16369 GCEIQYIVEKCDVNVGEWTMCTPTTGINKTNIEVEKLEKHEYNFRICAINKAGVGEHA 16428  
QY 290 ----- 289  
Db 16429 DVPQPIIVEKLEAPDIDLDELKIIINIRAGGSLRLFPVPIKGRPTPEVKWGVKVDGEIRD 16488  
QY 290 ----- 289  
Db 16489 AAIIIDVTSSFTSLVDNVRVDSKYTLTLENSSGTKSAFVTVRVLDTSPVNLKVTEI 16548  
QY 290 ----- 289  
Db 16549 TKDSVSIWBPPLLDGGSKIKNIYIVEKREATRKSAAVVTNCHKNWKIDQLQEGCYSYF 16608  
QY 290 ----- 289  
Db 16609 RVTAEYIGLPAQTADPIKVAEVPQPGKITVDDVTRNSVSLSWTKPEHGGSKIIQY 16668

QY	290	-----TNVT-----	293
Db	16669	IVEMQAKHSEKWCARVKSQAVITNLTOGEYLFRRVAVNEKGRSDPRSLAPIVAKD	16728
QY	294	-----	293
Db	16729	LVIEPDVKPAPSSYSVQVQDLKMEVPIISGRPKPTITWTGDLPLKQTTTRINVTDSLOLT	16788
QY	294	-----YSITTA-----	299
Db	16789	TLSEIKETHKDDGGQGITVANVVGQKTASIEIVTLDKPPPKGPKVFPDDVSAESITLSMN	16848
QY	300	-----	299
Db	16849	PPLYTGGCQITNVIQKRDTTTTVMDVVSATVARTTLKTKTGTQYQFRIFAENRYCQ	16908
QY	300	-----	299
Db	16909	SPALESDPIVAQYPYKEPGPGTPTATAISKDSMVIQWHEPVNNGSPVIGYHLERKERN	16968
QY	300	-----	299
Db	16969	SILTKVNTKTIIDTQFKAQNLBEGIEYFRVVAENIVGVGASKNSKSECYVARDPCDPPG	17028
QY	300	-----	299
Db	17029	TPEPIWKRNEITLQWTKPVYDGGSMITGYIVEKRDLPDGRWMKASFTNVIETQFTVSGL	17088
QY	300	-----	299
Db	17089	TEDQYFRVIAKNAAGAISKPSDSTGPITAKDEVLPRIKDEVELPRISMDPKFRDTIIVNAGETFRL	17148
QY	300	-----	299
Db	17149	EADVHGKPLPTIEWLRGDKEIEESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAG	17208
QY	300	-----	299
Db	17209	SKSPFVNKVLDRPGPEPGVQVGTSEKCSLTWSPPLQDGGSDISHYVVEKRETSRLA	17268
QY	300	-----	299
Db	17269	WTVASEVVTNSLKVTKLLEGNEYFRVIMAVNKYGVGBLESAPVLMKNPFLPFPKPSL	17328
QY	300	-----ARRV-----	303
Db	17329	EVTNIAKDSMTVCWNRPSDGGSEIIGYIVEKRDSEIRWIKCNKRITDRLRLVGTGLTE	17388
QY	304	-----STSTIAYR-----	311
Db	17389	DHEYFRVSAENAAGVGPSPATVYKACDPVKPGPTNAHIVDTTKNSITLAWGKPIY	17448
QY	312	-----	311
Db	17449	DGSEILGYVVEICKADEEWOIVTPQGLRVTRFEISKLTQHOYKIRVCALNKVGLGE	17508
QY	312	-----	311
Db	17509	ATSPGTVKPEDKLEAPELDLDELKGIIVRAGSARIHIPFKGRPMPEITWSREGEF	17568
QY	312	-----	311
Db	17569	TDKVQIEGVNVTQLSIDNCDRDAGKYILKLENSGSKSAFVTVKVLDTGPPQNLAVK	17628
QY	312	-----	311
Db	17629	EVRKDSAPLWEPPIIDGAKVKVVIDKRESTRKAYANVSSKSKTSFKVENITEGAIY	17688
QY	312	-----	311
Db	17689	YFRVMAENEFGVGVFVETDAVKAAPSPPGKVTLTVDVTSASLWKEPHEHGGSRVL	17748
QY	312	-----	311

Db	17749	GYVVMQPKGTEKWSIVAESKVCNAVVTGLSSGOEYQFRVKAAYNEKSGKSDPRVLGVPVIA	17808
QY	312	-----	311
Db	17809	KDLATQPSLKLFFNTYSIQAGEDLKIEIPVIGRPRPNISWVKDGEPLKQTTTRVNVVEETAT	17868
QY	312	-----	311
Db	17869	STVLHIKEGNKDDPGKYTVTATNSAGTATENLSVILEKPGPVGPVRFDSVSADPVVIS	17928
QY	312	-----	311
Db	17929	WEPPAYTGGCQISNYIVEKRDTTTTTHMVSATVARTTIKITKLTGTQYQFRIFAENRY	17988
QY	312	-----	311
Db	17989	GKSAPLDSKAVIVQYPFKEPGPGTPTFVTSISKQMLVQWHEPVNDGGTKIIGYHLEQKE	18048
QY	312	-----	311
Db	18049	KNSILWKLANKTPIQDTFKTTGLDLEGEYEFKVSANIVGIGKPSKVSECFVARDPCDP	18108
QY	312	-----PDSSPMKS-----	319
Db	18109	PGRPEAIVITRNNTVLKWKPAYDGGSKIITGYIVEKKDLDPGRWMKASFTNVLETEFTVS	18168
QY	320	-----	319
Db	18169	GLVEDQYFRVIAARNAAGNFSEPSDSSGAIARTARDEIDAPNASLDPKYKDVIVVHAGETP	18228
QY	320	-----	319
Db	18229	VLEADIRGKPIPDVVWSKDGELEETAARMEIKSTIOKTTILVVKDCIRTDGGQYILKLSN	18288
QY	320	-----	319
Db	18289	VGGTKSIPITVKVLDRPGSPGPKVGTVAEKCYLAWNPLODGGGANISHYIEKEETS	18348
QY	320	-----IMA-----	322
Db	18349	RLSWQTVSTEQVQALNYKVTKLLFCNEYIFRVMVKNKYGIGEPLESGPVTACNPKYPPGP	18408
QY	323	-----TOLR-----	326
Db	18409	STPEVSAITKDSMVVTVARPVDDGGTEIEGYILEKRDKEGVRTKCNKTLTDLRLVGT	18468
QY	327	-----	326
Db	18469	LTEGHSYFRVAAENAAGVGESEPSVIFYRACDALYPPGPPSNPKVTDTSRSSVSLAWSK	18528
QY	327	-----	326
Db	18529	PIYDGGAPKGYVVEVKEAAADEWTTCTPTGLQKQFTVTKLKENTYFNFRICAINSEG	18588
QY	327	-----	326
Db	18589	VGEPATLPGSVVAQERIEPPEILDADLRKVVLVRSATLRLFTTIGRPEPEVKWEKAE	18648
QY	327	-----	326
Db	18649	GILTDRAQIEVTSFTMLVIDNVTRFDSGRYNTLLENNSGSKTAFVNVRLDSPSAPVNL	18708
QY	327	-----	326
Db	18709	TIREVKKDSVTLSEWEPPLIDGGAKITNYIVEKRETRKAYATITNNCTKTTPRIENLOEG	18768
QY	327	-----	326
Db	18769	CSYFRVLASNEYGIGLPAETTEPVKVSBPPLPPGRVTLVDVTRNTATIKWEKPESDGGS	18828
QY	327	-----	326



Db	18829	KITGVVEMQTKSEKMSCTQVTKLEATISGLTAGEEYVFRVAANEKGRSDPRLGVP	18888
Qy	327	-----	326
Db	18889	VIARDIEIKPSVELPFTFNKAREQLKIDVPFKGRQATVNRKDGQTLKELTRVNVSS	18948
Qy	327	-----	326
Db	18949	SKVTSLSKEASKEDVGTYELCVNSAGSITVPTIIVLDRPGPGPIRIDEVSCDSIT	19008
Qy	327	-----	326
Db	19009	ISWNPPEYDGGCQISNIYVEKETTSTTHIVSQAVARTSIKIVRLTTGSEYQFVCAEN	19068
Qy	327	-----	326
Db	19069	RYGKSSYSESSAVAEYFPSPGPGTPKVVHATKSTMLTWQVPVNDGGSRVIGHLEY	19128
Qy	327	-----	326
Db	19129	KERSILMSKANKILIAQTQVKVGLDEGLMYEYRVAENIAGICKSKSCPEVPARDPC	19188
Qy	327	-----	326
Db	19189	DPGQPEVTNITRKSLSLWSPHYDGGAKITGYIVERRELDPGRWLKCNNTNIQETFE	19248
Qy	327	-----	326
Db	19249	VTELTEDQRYEYFRVARNADSBSBSESTGPIIVKDDVEPPRVNMDVKFRDVIWKAGE	19308
Qy	327	-----	326
Db	19309	VLKINADIAGRPLPVISWAKDIEIEERARTEIISTDNHTLLTVKDCIRRDGTQVVLTK	19368
Qy	327	-----	326
Db	19369	NVAGTRSVANCKVLDPKPPAGPLEINGLTAEKCSLSWGRPOEDGGADIDYHKKRET	19428
Qy	327	-----	326
Db	19429	SHLAWTICEGELQMTCKVTLLKNGEYIFRVTGNVKGVEPLBSVAIKALDPPTVPS	19488
Qy	327	-----	326
Db	19489	PTSLEITSVKESMTLCSRPESDGSGISGYIIRERKNSLRVVRNKKFVYDLRVKST	19548
Qy	327	-----	326
Db	19549	GLREGCEYRYAENAGLSLPSSETSPLIRABDPVFLPSPSPKIVDSGKTTITIAW	19608
Qy	327	-----	326
Db	19609	KLPFDGGAPITGYTVEYKSDTDWKTISIQLRGTEYITISGLTTGAEYVFRVKSUNKVA	19668
Qy	327	-----	326
Db	19669	SDPSSDPQIAKEREEREPFLDIDSEMRKTLIVKAGASFTMTVPRGRPVENVLWSKPD	19728
Qy	327	DLAT-----	330
Db	19729	DLRTRAYVDTDSRTSLTIENARNRDSKTYTLTIQNLVLSAASLTLVVKVLOTGPPPTNIT	19788
Qy	331	-----	330
Db	19789	VQDVTKESAVLSWDVPENDGGAPVKNYHIKREASKKAWSVTNNCNELSYKVTNLQEGA	19848
Qy	331	-----	330
Db	19849	IYFRVSGENBFGVIGPAETKEGVKITEKSPPEKLGVTISKDSVSLTWLKPHEHGGSR	19908
Qy	331	-----	332
Db	19909	IVHYVVEALEKQKNNVKAIVAKSTHVVVSGLRENSYFFRVAENQAGLSDPRELLPV	19968
Qy	333	-----	332
Db	19969	LIKEQLEPEIDMKNPFSTHTVYVAGSNLKVDIPIFGKPLPKVTLSDRGVPLKATMRFT	20028
Qy	333	-----	332
Db	20029	EITAEINTLNKESVTADAGRYEITANSSGTTKAFINIVVLDPRGPPTGPVVISDITEE	20088
Qy	333	-----	332
Db	20089	SVTLKWBEPKYDGSQVNTYILLKRETSATVMTVEVSATVARTMMKMKLTTGEEYQPRIK	20148
Qy	333	-----	335
Db	20149	AENRFGISDHIDSACVTVKLPYTPPGPSTPWVNTVRESITVGMHEPVSNGSAVGYH	20208
Qy	336	-----	335
Db	20209	LEMKDRNSILMQKANKLVIRTHFKVTTISAGLIYBFRVYAENAGVGKPSHPSEPVLA	20268
Qy	336	-----	335
Db	20269	DACEPPNRVITDISKNSVSLWQOPADGGSKITGYIVERREDLPDGRWTKASFTNVTE	20328
Qy	336	-----	339
Db	20329	QFTISGLTQNSQYBFRVARNAGSISNPSEVVGPIITCIDSYGGPVIDLPLEYTEVVKYR	20388
Qy	340	-----	339
Db	20389	AGTSVKLRAGISGKPAFTIEWYKDDKELQTNALVCVENTTDLASILIKDARLNSGCYEL	20448
Qy	340	-----	339
Db	20449	KLRNAMESATIRVQILDKPGPGPIEFTVTAEKITLLWRPADDGGAKITHYIVBK	20508
Qy	340	-----	339
Db	20509	RETSRVVMSVSEHLEECIITTKIIGNEYIFRVAVNKYGIGEPLESDSVAKNAFVT	20568
Qy	340	-----	339
Db	20569	PGPGIPEVTKITKNSMTVVMRSPIADGSGDISGYFLEKRDKSLGWFVKLKETIRDRQ	20628
Qy	340	-----	347
Db	20629	KVTGLTENSQYQYRVCANAGQGFSEPFYKAADPIDPPGPAKIRIADSTKSSITL	20688
Qy	348	-----	347
Db	20689	GWSKPVYDGSASVTGYVVEIRQGEEBEWTTVSTKGEVTRTEYVYVNLKPGVNYFRVSAV	20748
Qy	348	-----	347
Db	20749	NCAGQGEPIEMNEPVQAKIDLEAPEIDLVALRTSVIAKAGEDVQVLIPFKRPPPTVTM	20808
Qy	348	-----	351
Db	20809	RKDEKNLGSADARYSIENTDSSLLTIPQVTRNDTGKYLITTIENGVEGPKSTSVKVLDT	20868
Qy	352	-----	351
Db	20869	PAAQKQLQVHVSRTVTLWDPLIDGGSPIINNVIEKRDTAKRTWSVSHKCSSTSPK	20928
Qy	352	-----	351
Db	20929	LIDLSEKTPFPFRVLAENEIGIGEPCEPTEPVKAAEVPAPIRDLMSKDKSTKSVLSWK	20988
Qy	352	-----	351
Db	20989	PDFDGGSVITEYVVERKKGQWTHAGISKTCBIEVSQLEQSVLEFPRVPAKNEKGLSD	21048

QY	352	-----	351
Db	21049	PVTIGPITVKELIITPEVLDSDIPGAQVTVRIGHNVHLELPYKGPKPSPISWLKDLPLK	21108
QY	352	-----	351
Db	21109	ESEFVRPSKTENKITLSIKNAKKEHGKQYTVILDNAVCRITAVPITVITLPPSPKPGPIR	21168
QY	352	-----	351
Db	21169	FDEIKADSVILSWDPEDNGGGEITCYISIEKRETSQTNWKMVCSSVARTTFKVPNLVKDA	21228
QY	352	-----	351
Db	21229	EYQFVRABNRYGVSQPLVSSIIIVAKHQFRIIPGPPGKPVINYNTSGMSLTWDAPVYDGG	21288
QY	352	-----	351
Db	21289	SEVTGFHVEKERNILWQVNTSPISGREYRATGLVEGLDYQFRVVAENSAGLSSPSDP	21348
QY	352	-----	351
Db	21349	SKFTLAVSPVDPGTPYIDVTRETITLKNWPLRLDGGSKIVGYSIEKRGNERWRCNF	21408
QY	352	-----	351
Db	21409	TDVSECQYTVTGLSPGRYEFRIIARNAGVTISPPSQSSGIIMTRDENVPPIVEFGPEYF	21468
QY	352	-----	351
Db	21469	DGLIKSGESLRIKALVQGRPVRVTFWFDGVEIEKRMMEITNVLSGTSLFVRDATRDH	21528
QY	352	-----	351
Db	21529	RGVYVVEAKNAGSAAEIKVKVQDTPGKVVGPIRFTNITGEXMTLWDAPLNDGCAPIT	21588
QY	352	-----AVSEF-----	356
Db	21589	HYIIKRETSRLAWALIEDKCEAQSYTAIKLINGNEYQFRVSAVNKFGVGRPLDSDPVVA	21648
QY	357	-----	356
Db	21649	QIQYTVDPDAPGIPESNITGNSITLTWARPESDGGSEIQOYILEREKKSTRWVKVSKR	21708
QY	357	-----	356
Db	21709	PISETRFKVTGLTEGNEYEFHVAENAAAGVPASGISRLIKCREPNPVPPTWVKVTD	21768
QY	357	-----	356
Db	21769	SKTTVSLEWSKVPDGMIEIIGYIEMCKTDLGDWHKVNAAECVTRYTVTDLQAGEYK	21828
QY	357	-----MKNTHVL-----	363
Db	21829	FRVSAINGAGKDSCEVTGTIKAVDRLTAPELIDANFKQTHVVRAGASIRLFIAYQGRP	21888
QY	364	-----	363
Db	21889	TPTAVMSKPSNLSLRADIIHTDSFSTLTIVENCNRNDAGKYTLTVENNSSGKSIITFTVKV	21948
QY	364	-----	363
Db	21949	LDTFGPPGPIITFKDVTGRSATLMDAPLLDGGARIIHHYVVEKREASRRSQVISEKCTRQ	22008
QY	364	-----	363
Db	22009	IFKNDLAEQVPYPRVSAVNEYGVGEPEPIVATEQPAPPRRLDVDTSKSSAVLA	22068
QY	364	-----	363
Db	22069	WLKPDHGGSRITGYLLEMRQKSDLWVEAGHTKQLTFTVERLVEKTEYEFVRKAKNDAG	22128
QY	364	-----	363

```
Db 23209 LPRIARQNEKGVSEPRETTTAVTVQDLRLVPLTIDLSTMPQKTIHVPAGRPVELVPIAGR 23268
Qy 382 ----- 381
Db 23269 PPPAASWFFAGSKLRESERVTVETHTKVAKLFTIRETTTIDTGEYTLKLNVTGTTSETIK 23328
Qy 382 ----- 381
Db 23329 VIILDKPGPTGPIKIDIDATSITISWEPPELDGGAPLSGVYVBOADHRPGLPVSSES 23388
Qy 382 ----- 381
Db 23389 VTRSTFKRLTEGNEYVFRVAATNRFGLGSVLOSEVIECRSSIRIPGPPTLQIFDVSR 23448
Qy 382 ----- 381
Db 23449 DGMILTWPPEDDGSGVTGYIVERKEVRADRWVRNVKVPVMTTRYRSTGLTEGLEYEHR 23508
Qy 382 ----- 381
Db 23509 VTAINARGSGKPSRPSKPIVAMPDPIAPPCKPQNPRVTDTRTSVSLAWSVPEDEGSKVT 23568
Qy 382 ----- 381
Db 23569 GYLIEQKVQDQHEWTKCNTPTPKIRBYTLTHLPQGAERYFRVLACNAGGPGEPAPVPGTV 23628
Qy 382 ----- 389
Db 23629 KVTEMLPYDPVELDERYQEGIFVRQGVIRLTPIPIKGKPPICKWTKEGQDISKRAMIAT 23688
Qy 390 ----- 392
Db 23689 SETHTELVIKEADRGDSGTDLVLEKCGKKAUYIKVRVIGSPNSPEGLBYDDIQVRSV 23748
Qy 393 ----- 392
Db 23749 RVSWRPADGGADILGILERREVPAKAWTYTIDSRVRGTSVLVKGLENVEYHFRVSAE 23808
Qy 393 ----- 392
Db 23809 NQFISGKPLKSEEPVTKPLNPPPPNPPPEVLDVTKSSVLSWSRPKDGGSRVTGY 23868
Qy 393 ----- 406
Db 23869 IERKETSTDKVVRHNKTOITMTYTGLVPAEYQFRIIAQNVDVGLSETSPASEPVVCK 23928
Qy 407 ----- 409
Db 23929 DPFKPSQPELEILSISKDSVTLQWKEPCDGGKEILGYVWYVRQSGDSAWKSNKERI 23988
Qy 410 ----- 409
Db 23989 KDKQFTIGGLEATEYEFVRFAENETGLSRPRRTAMSIKTKLTSGEAPGIRKEMKDVTK 24048
Qy 410 ----- 409
Db 24049 LGEEAQLSCQVGRPLDKWYRFGKELIQSKYKMSDGRTHLTVMTEQEDEGVYTC 24108
Qy 410 ----- 409
Db 24109 IATNEVGEVETSKLLLOATQFHPGYPLKBYKYGAVGSTLRLHVMYIGRPVPAWTFHG 24168
Qy 410 ----- 409
Db 24169 QKLLQNSENITIENTEYTHLVKMKVQRKTHAGKYKVQLSNVFGTVDAILDVEIQDKPDK 24228
Qy 410 ----- 409
Db 24229 PTGPVIEALLKNSAVISWKPADDGGSWITNYVVEKCEAKEGAEWQLVSSAISVTTCRI 24288
Qy 410 ----- 417
Db 24289 VNLTENAGYFRVSAQNTFGISDPLEVSSVVIKSPFEKPGAPKPTITAVTKDCSVVAV 24348
```

---

```
Qy 418 ----- 417
Db 24349 KPPASDGAKIRNYYLEKREKKQNKWISVTTEEIRETVFSVKNLIEGLEYEPRVKCNILG 24408
Qy 418 ----- 417
Db 24409 GESEWSEISEBITPKSDVPIQAPHPKBEELNLRVYQSNATLVCKVTHPKPIVKYRQG 24468
Qy 418 ----- 417
Db 24469 KEIIADGLKYRIOBFKGGYHQLIIASVTDDATVYQVRATNQGGSVSGTASILEVEVPAKI 24528
Qy 418 ----- 417
Db 24529 HLPKTLGEMGAVHALRGEVWSIKIPFGKPDVITWQKQDLIDNNGHYQVIVTRSF7SL 24588
Qy 418 ----- 417
Db 24589 VFPNGVERKDAGFYVCAKNRFGIDQKTVELDVADVPDPGRGVKVDASRDSVNI7WTEP 24648
Qy 418 ----- 417
Db 24649 ASDGSKITNYIVEKCATTAERLVRVGOARETRYTVINLFGKTSYQFPRVIAENKFGLSKP 24708
Qy 418 ----- 417
Db 24709 SEPSEPTITKEDKTRAMNYDEBDBREVSMTKASHSTKELYEKYMAEDLGRCEFGIV 24768
Qy 418 ----- 417
Db 24769 HRCVETSSKTYMAKFKVKGTDQVLVKEKISILNIARHNLHLHESFESMEELVMPIFE 24828
Qy 418 ----- 417
Db 24829 FISGLDIFERINTSAPELNEREIVSVYHVQCEALQFLHSHNIGHFDIRPENIIYQTRSS 24888
Qy 418 ----- 418
Db 24889 TIKIIFQARQLKPGDNFLLFTAPEYVAPEVHQHDVSTATDMWSLGLTVLVLLSGIN 24948
Qy 419 ----- 425
Db 24949 PFLAETNQIIEINMNAEYTFDEBAPKEISIEANMDFVDRLLVKKERKSRMTASEALQHPWL 25008
Qy 426 ----- 425
Db 25009 KQKIERVSTKVIRTLKHRRYHYHTLIKDLNVMVSAARISCGAIRSQKGVSAKVAVASI 25068
Qy 426 ----- 433
Db 25069 EIGPVSGQIMHAVGEEGHVKYCKIENYDQSTQVTWYFGVRLSEKYEITVEDGVAI 25128
Qy 434 ----- 433
Db 25129 LYVKDITKLDGTYRCVNDYDGBSDSYAEFLVKGREVDYDCCRRTMKKIKRRTDTMRL 25188
Qy 434 ----- 433
Db 25189 LERPPEFTLPLYNKTAUVGENVRFGVITVHPEPHVTWYKSGQIKPGDNKKYTFESDK 25248
Qy 434 ----- 433
Db 25249 GLYQLTINSVTDDDAEYTVVARNKYGBSDCKAKLTVLHPPTDSTLRLPMFKRLLANAE 25308
Qy 434 ----- 433
Db 25309 COEQSVCFEIRVSGIPPPPLKWEKGQPLSLGPNIEIIEHGLDYALHIRD7LTPEDTGY 25368
Qy 434 ----- 433
Db 25369 YRV7ATN7AGSTSCQAH7QVERLYKKOE7FKSBEHERHVQKIDK7L7RMAEILSGTESV 25428
```

Qy 434 ----- 433  
Db 25429 PLTQVAKALREAAVLKPAVSTKTVKGFRLEIEBKEERKLMPYDVPBRKYQTTI 25488  
Qy 434 ----- 433  
Db 25489 EBDQRIKQFVPM5DMKWYKIRDQYEMPGKLDREVQKPRKRLRLSRWEPQFVMPPLRITD 25548  
Qy 434 ----- 433  
Db 25549 QYRPKWRIPKLSQDLEIVRPARRTPSPDYFYFRRRRSLGDISDELLLPIDDLAM 25608  
Qy 434 -----SLR----- 436  
Db 25609 KRTEBERLRLEBELELGF5ASPPSPHPFELSSRLYSSPOAHVKVETZRKNFRSTYHI 25668  
Qy 437 ----- 436  
Db 25669 PTKAEASTSYAELRRHQAAYRQPKORQIMAREDEELLRPVTTTQHLSEYKSELDFM 25728  
Qy 437 -----SPTVNL----- 443  
Db 25729 SKEEKSRKSRQREVTEITEIEEYELSKHAQRESSSASRLRRRRSLSPYIELMRP 25788  
Qy 444 -----TP----- 445  
Db 25789 VSELIRSPQAEYEDDTERRSPTPERTPRSPSPVSSRSLSRPFSGARDFISRYES 25848  
Qy 446 ----- 445  
Db 25849 MKAALKTKTSEKYEVL5QOFTLHDAPRITLMRSHRVPQGNQTRFILNVQSKPTAEV 25908  
Qy 446 ----- 445  
Db 25909 KWHNGVELOESSKIHYNTSGVLTLEILDCHDDSGTYRAVCTNYKGEASDYALDVTG 25968  
Qy 446 ----- 445  
Db 25969 GDYTYASQRDEVRSPPELTRTEAVAVSPFKTSEMEASSSVREYK5QMTETBSL 26028  
Qy 446 ----- 445  
Db 26029 SSYHSASABMSAALKEK5LEK5TRKIKTTLAARILTKPR5MTVVEG5AR5P5C5DTD 26088  
Qy 446 -----PEHRAV----- 445  
Db 26089 GEPVPTVWLKQVLT5ARHQVTTTKYKSTFEISSVQASDEGNYSVVVENSEGKQEA5 26148  
Qy 446 -----PEHRAV----- 452  
Db 26149 FTLTIQARVTEKAVT5PPRVK5PEPRVKSPEAVK5PPE5HPKAV5PTETKPTP 26208  
Qy 453 ----- 452  
Db 26209 IEKVQHLFVSAPPK1TQFLKABASKEIAKLCTVWESSVLRAKEVTWKDGKLGKENGHFQ 26268  
Qy 453 -----NLSTN----- 458  
Db 26269 PHYSADGTVELKINNLTESDQGEYVEISGEGGTSKTNLQPMQOAFK5IHEK5K1SETK 26328  
Qy 459 -----SLW 461  
Db 26329 KSDQKTST5TRKTEPKAPFIS5KPVITGLQD5TVSSD5VAKFAVKATG5PR5TAIW 26388  
Qy 462 ----- 461  
Db 26389 TKOGKAITQGGKYL5EDKGGFFLEIHK5TDSGLYTC5TVKNSAG5VSS5CKL5TIKA5K 26448  
Qy 462 ----- 461  
Db 26449 DTEAQK5V5TK5BITPQKAVVQ5EISQKALR5EEIKM5EAK5Q5EKLAK5EASKVL5S 26508  
Qy 462 ----- 461

Db 26509 EEVKKSAA5LSK5IVH5EITK5QASE5EVRTHABIKAF5TQMSINEQRLVLKANIAGA 26568  
Qy 462 -----W----- 462  
Db 26569 TDVKWVLNGVELTNE5EYRVG5GSDQTLTIKQASHRDEGILTCISKYKEIVKQCYDLT 26628  
Qy 463 -----WLQ 465  
Db 26629 LSKELSDAPAFISQPR5QNR5QIN5EQNVLFTR5EISG5P5EIEW5FK 26672

RESULT 3  
S41309

C:Species: cyclosporin fungus  
C:Species: Tolypocladium inflatum (cyclosporin fungus)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: S45487; S41309  
R:Weber, G.; Schoergerdorfer, K.; Schneider-Scherzer, E.; Leitner, E.  
C:Title: The peptide synthetase catalyzing cyclosporine production in Tolypocladium nivei  
A:Reference number: S45487; MUID:95094306; PMID:8001164  
A:Accession: S45487  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-15281 <WEB>  
A:Cross-references: UNIPROT:Q09164; EMBL:Z28383; NID:g440168; PIDN:CAA82227.1; PID:g44401  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993  
A:Note: only a part of the translation is shown  
A:Note: the source is designated as Tolypocladium inflatum  
C:Genetics:  
A:Genetic code: SGC3  
C:Superfamily: cyclosporin fungus cyclosporin synthetase; acetate-CoA ligase homology; a  
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F:532-1008/Domain: acetate-CoA ligase homology <ACLI>  
F:1029-1096/Domain: acyl carrier protein homology <ACP1>  
F:1618-2069/Domain: acetate-CoA ligase homology <ACLI>  
F:2527-2594/Domain: acyl carrier protein homology <ACP2>  
F:3115-3553/Domain: acetate-CoA ligase homology <ACLI>  
F:4014-4081/Domain: acyl carrier protein homology <ACP3>  
F:4601-5057/Domain: acetate-CoA ligase homology <ACLI>  
F:5506-5573/Domain: acyl carrier protein homology <ACP4>  
F:6094-6546/Domain: acetate-CoA ligase homology <ACLI>  
F:7003-7070/Domain: acyl carrier protein homology <ACP5>  
F:7591-8042/Domain: acetate-CoA ligase homology <ACP6>  
F:8063-8130/Domain: acyl carrier protein homology <ACP7>  
F:8652-9157/Domain: acetate-CoA ligase homology <ACP8>  
F:9558-9625/Domain: acyl carrier protein homology <ACP9>  
F:10146-10586/Domain: acetate-CoA ligase homology <ACP10>  
F:11055-11122/Domain: acyl carrier protein homology <ACP11>  
F:11635-12106/Domain: acetate-CoA ligase homology <ACP12>  
F:12127-12194/Domain: acyl carrier protein homology <ACP13>  
F:12715-13159/Domain: acetate-CoA ligase homology <ACP14>  
F:13623-13690/Domain: acyl carrier protein homology <ACP15>  
F:14213-14676/Domain: acetate-CoA ligase homology <ACP16>  
F:14698-14765/Domain: acyl carrier protein homology <ACP17>  
F:1060-2558,4045,5537,7034,8094,9589,11086,12158,13654,14729/Binding site: phosphopantet:

Query Match 23.5%; Score 575; DB 2; Length 15281;  
Best Local Similarity 1.7%; Pred. No. 0.047;  
Matches 254; Conservative 78; Mismatches 128; Indels 14673; Gaps 82;  
Qy 1 MG----- 2  
Db 1 MGAIGQDMAYDRLANP5R5SSIS5NRY5PE5VQ5FAQ5RL5FLH5QL5K5GAS5DIT5PAIR 60  
Qy 3 ----- 2  
Db 61 LRGLDITDALNA5RALTQ5RHETL5TTF5KEQ5GVQ5VHV5ASGL5RGL5RIV5DASS5RDLAQ 120  
Qy 3 ----- 2  
Db 121 LLABEQTMKFDL5EPAP5RWVALL5KVAED5HHIL5IVVHHI5SDRS5LDIT5QOEL5ELYTAA 180

```
Qy 3 ----- 2
Db 181 SQKSIACPLGPIPIQYRDLTTWQNOQDEQVAEQERQLGYWIEQDNNTPAELLTELPRP 240
Qy 3 ----- 8
Db 241 AIPSGETKISFOIDGSHKELLAPCRSQQVTAVALLAAPRAVHFRITGAEDATIGAPV 300
Qy 9 ----- 12
Db 301 ANRDRPELENVVAPLATIQCMRVLDDEDTFESVLRQIMSVWTEAHANRDVPPERIYSAL 360
Qy 13 ----- 17
Db 361 LPGSTDSRHLPLVQLMFLALHPAQDTGRARWGFLBAETTLQSAAPTRFDWMEMHLPBGDDRFD 420
Qy 18 ----- 17
Db 421 ANVLFTGLPDAEAIKRSVVSIFREVLRRGISEPAVHVKTMTPLTDGLAAIRDMLLDIGTT 480
Qy 18 ----- 17
Db 481 DYPREASVDMFQEQVALNPATAVADASSRSLSYSELDHKSQDLAAMLRRLQKPELTIG 540
Qy 18 ----- 17
Db 541 VLSPPSCETWVSFLGILKAHLAYPLDINVPLARIESILSAVDGHLVLLGSLNVPQKVD 600
Qy 18 ----- 17
Db 601 VPDVELLRISDALNGSQVNLGAGKOATAKPSATDLAYVIFTSSTGPKPKGMIEHRGIVR 660
Qy 18 ----- 17
Db 661 LVKGTNIISPAQAAPTAHLANIAFDLSTWEIYTPILNGGTLVCIHVSHTLDSKALEAVP 720
Qy 18 ----- 23
Db 721 TKEGIRVAFLAPALIKOCLADRPALFAGLDSLYAIGDRFDRRDALHAKSLVHGYNAYG 780
Qy 24 ----- 23
Db 781 PTENSVVSTIYSVSEASPFVTGVPVGRAISNGAYVMDQDQQLVSPGVGMLVSGDLA 840
Qy 24 ----- 23
Db 841 RGYTDSALDKNRFVVVQIDGESIRGYRTGDRARYSLKGGQIEFFGRMDQVKIRGHRIEP 900
Qy 24 ----- 23
Db 901 AEVEHALLNSQVRDAVWIRRQEEBPAMIAFVTTQGTLPDLVNLNGHVPDNGSK 960
Qy 24 ----- 23
Db 961 NDQPAVHVESLRRRLQMLLPSYMMPARIVVLDHLPLNPGKVDKALGOSAKTVQSKL 1020
Qy 24 ----- 23
Db 1021 VSQVAPRNEIEAVLCEEYRSVLGVGEVIGITNPFDLGHSHTAMKLAARISQRLDIQASV 1080
Qy 24 ----- 23
Db 1081 ATVPEQPLADLAATIQRGSLYSVIPTEYTGVPVQSPAGQLWFLQNLGTGASWYNVM 1140
Qy 24 ----- 23
Db 1141 LTVRLRGLDVLGALTALLALEKREHETLRTFFEBRDGVMQVHSSLMGELRLIDISEKS 1200
Qy 24 ----- 23
Db 1201 GTAAHEALMKEQSTRFDLITREPGWRVALLKLADHHIFSIVMHHIVSDGWSLDDLRLHELGG 1260

Qy 24 ----- 23
Db 1261 LYSAAARGQDPLSRLEPLPIQYRDFAVWQKDSQKAAHQRLEYWTKQLADSTPAELLT 1320
Qy 24 ----- 23
Db 1321 DFPSPSILSGKAGKVPVAIBESLYDTLQVFSRTHQVTSFAVLLAAFRAAHFLTGSDDNT 1380
Qy 24 ----- 23
Db 1381 IGVPSANRNRPELENVIGFFVNTQICIRITIDENDNFESLVRQVSRSTTTAAQNDQDVPEQ 1440
Qy 24 ----- 23
Db 1441 VSSIMPSSSRDASRNPLVQLMFLAHGQDLFKIQLEGTSEBEVITEBTVRFDIEPHLYQ 1500
Qy 24 ----- 23
Db 1501 GASKLSGDIIPAADLPEAETIRGVSVFQEVLRGLQOPQTPIMTMTDGIPELERMGL 1560
Qy 24 ----- 23
Db 1561 LHMVKTDYPRNMSVVDVFPQQVRLSABATAVIDSSSRMSYAELOQRSQVAAWLRQRQLP 1620
Qy 24 ----- 23
Db 1621 AETFVAVLAPRSCBAVIALFGILKAGHAYLPDYNVPAARLRAILAEVKGKLVLLGAGE 1680
Qy 24 ----- 23
Db 1681 PSPEGQSPESIVIRIADATSPAGHASLRDGSKSKTAGSLAVIIFTSSTGPKPKGMIEHR 1740
Qy 24 ----- 23
Db 1741 GVLRLVKQTNILSSLPAPQTFERMAHMSNLAFDASIWEIFTALLNGGSLVCIDRFTILDAQ 1800
Qy 24 ----- 23
Db 1801 ALEALFUREHINTALPPPALLKQCLTDAATIKSLDLLLYVGGDRDLTADAAALAKLVKSE 1860
Qy 24 ----- 23
Db 1861 VYNAYGTENTVMTSLYIADTERFVNGVPIGRAVNSGVYVMDQNLQPLVGVNGELVV 1920
Qy 24 ----- 23
Db 1921 TGDGLARGYTNPALDSRDFVDVVIARGQLLRAYRTGDRARYRPKDQVEFFGCRMDHQVKVR 1980
Qy 24 ----- 28
Db 1981 GHRIELAEVEHALLSSAGVHDVAVVNSQEDNQGVEMVAFITAQDNETLQEAQSSNQVOE 2040
Qy 29 ----- 28
Db 2041 WESHFETTAYADITADONTLGRDFTSWTSMYDGLIDKREMQEWLDDTMTFTLDGQAAG 2100
Qy 29 ----- 28
Db 2101 HVLEIGTGTGWLNLGQAGLKSYIGLEPSQSAQVFNKAAQTFFGLEGKAQVHVGTAWD 2160
Qy 29 ----- 28
Db 2161 TGRLSALSPDLIVINSVAQYPPSREYLAEVVEALVRIPGVRRIFFGDMRTYATHKDFLVA 2220
Qy 29 ----- 28
Db 2221 RAVHTNGSKVTRSKVQOEVARLEEBELLVDPAFFTSLKESLSEIEHVEILPKNMKVN 2280
Qy 29 ----- 28
Db 2281 NELSSRYGAVLHTRHNQNSRHSIHKINAESWIDPASSQMDROGLARLLKENKDAESIA 2340
Qy 29 ----- 28
```

Db	2341	VFNIPVSKTIVERHIAKSLADDDGDDTHSSIDGVAMISAAAREKASQCSPLDVHDLVQLA	2400
Qy	29	-----	28
Db	2401	EDAGPRVEVSWARORSONGALDVFFHHFOPTENESRALVDFPTDYKGOQARSLTNRPLQR	2460
Qy	29	-----	28
Db	2461	VESRRIEAQVREQLVLLPAYMIPARIIVVLQNMPLNTSGKVDKRLTLRAKTAARTPSS	2520
Qy	29	-----	NCKVMSKA 36
Db	2521	ELVAPRDSIEALICKEFKDVLGVEVGI TDNFFNVGHSILATKLAARLSROLNAQIAVKD	2580
Qy	37	LYNRP-----	41
Db	2581	IPDRPVIADLAATIOODTTEHNPILPTSYTGVEQSPAQRMLFQLDQNLVGATWYLMPPA	2640
Qy	42	-----	41
Db	2641	VLRGPLVVSALAAALLAEERHETLRTTFIEQEGIGHQVTHPPAPKELRVIDVSGEES	2700
Qy	42	-----	41
Db	2701	TIQILEKEQTTPFNLAGSEPGFRLALLKTGEDEHILSTVMHHAISDGWSVDIFQOEIGQF	2760
Qy	42	-----	43
Db	2761	YSAILRGHDPLAQIAPLSIQVRDFATWQRIQFQVAEHRRLQAYWTKQLADKNKPAELLTDF	2820
Qy	44	-----	49
Db	2821	KRPPLMSGRAGEIPVVVDGLIVEKLQDFCRIRQVTAFTVLLAAFPAAHYRMGTGEDATIG	2880
Qy	50	-----	49
Db	2881	TPIANRNRPELEGLIGFVNTQCMRITVDVEDSFETLVHQVRETTAAAHANQDVPFEQIV	2940
Qy	50	-----	69
Db	2941	SNILPGSSDTSRNPVLQMLPALHSQONLGKVRLEGIEBIIISIAETTRFDIEFHLVQSAE	3000
Qy	70	-----	69
Db	3001	RLNGSIVVAADLVFPETIQSVITIFQGILOKGLGEPDMPVASMALDGGLESIRSTGLLHP	3060
Qy	70	-----	69
Db	3061	QOQDYPCDASVVQIFKQOVVNPVDVIAVRDESTRLSYADLDRKSDQVACWLSRRGIAPET	3120
Qy	70	-----	69
Db	3121	FVAILAPRSCETIVAILGLVKANLAYLPDVNVPASRLAILEILSEVSGSMLVLGAETPIP	3180
Qy	70	-----	69
Db	3181	EGMAEATIRITEILADAKTDDINGLAASQPTAASLAVIIFTSGSTGRPKGVVMVHRGIV	3240
Qy	70	-----	69
Db	3241	RLTKQTNITSKLPSFHHMAHISNLAFDASVWEVFTLLNGGTLCIDYFTLLLESTALEKV	3300
Qy	70	-----	69
Db	3301	FFDQVNVALLPALLKQCLDNLSPALVKLSVLVIGGDRLDASDAKARGLVQTAQFNAY	3360
Qy	70	-----	80
Db	3361	GPTENTVMSTIPIAEDPPFINGVPIGHAVSNSGAFVMDQNOQITPPGAGBELIVTGDGLA	3420
Qy	81	-----	80
Db	3421	RGYTTSSLNTGRFINVDIDGQVRAVRTGDRVVRPKDLQIEFFGRIDHQVKIRGHRIEP	3480
Qy	81	-----	80
Db	3481	AEVEYALLSHDLVTDAAVVTHSQBNQDLEMVGFVAARVADVREDESSNQVQWOTHFDSI	3540
Qy	81	-----	80
Db	3541	AYADITTIQQSLGRDFMSWTSMYDGSLLIKKSQMQEWLDDTWRSLLDSQPPGHVLEVTG	3600
Qy	81	-----	80
Db	3601	TGMVLENLREGGLQSVVGLRSPSATAFVNKAASPPGLEDRIRVEVGTATDIDRLGDD	3660
Qy	81	-----	85
Db	3661	LHAGLVVNSVAQYFPQDYLAQLVRDLTKVPGVERIPFGDMRSHAINRDLVARAVHAL	3720
Qy	86	-----	85
Db	3721	GDKATKAEIQREVVVRMEESEDELLVDPFAFTSLTTOVENIKHVEILPKRMRATNELSSYR	3780
Qy	86	-----	85
Db	3781	YAAVLHVNDLAKPAHKVSPGAWVDFAAATKMDRDALIRLLRGTKISDHTAIANIPNSKTIV	3840
Qy	86	-----	85
Db	3841	ERTICESVYDLGGDAKSDNDRVSWLSAARNAVKVASLSAIDLVDIAQAGRVEIISCAR	3900
Qy	86	-----	85
Db	3901	QWSQNGALDAVFHHLGPSQSSHVLIDFLTDHQRPREALTNHPLHRAQSRVRQIRER	3960
Qy	86	-----	97
Db	3961	LQTLTPAYMIPAQIMWLKPLNANGKVRKQLTQRAQTPVKAKQVSAPVAPRTEIERVL	4020
Qy	98	-----	97
Db	4021	COEPSVLVDIGIGIMENFFDLGGHSLMATKLAARISRRLETHSVSKSIFDHPRVCDLVI	4080
Qy	98	-----	110
Db	4081	VQOGSAPHDPVTSKTYGTPVQSPAQGRMLFLOLQNFEGATWYLMPLAVLRGAMNVHALT	4140
Qy	111	-----	110
Db	4141	AALLAERRHELLARTTFYEQNGVMQKNPVVTTETLRIIDLSNGDGDYLTPLKKEQTAPF	4200
Qy	111	-----	117
Db	4201	HLETEPCWRVALLRLPGDYILSVMMHIIISDGSVDVLFQELGQFYSTAVKGHDPLSQT	4260
Qy	118	-----	120
Db	4261	TPLPIHYRDPALWQCKPTQSEHERQLQYVWEQVDSAPABELLTDLRPSILSQAGEMS	4320
Qy	121	-----	123
Db	4321	VTIEGALYKNLEBFRCVRHVTSFVLLAALRAAHYRLTGSEDATIGTPIANRNRPELEQI	4380
Qy	124	-----	123
Db	4381	IGPFVNTQICIRITVNEDETPELSVQVRSATATAFAHQDVPFKEIVSTLLPGSRDASRNP	4440
Qy	124	-----	129
Db	4441	LVQLMFAVHSQKNLGELENHAHSEVVPTBITTRFDLEPHLFQODDKLEGSILYSTDLFE	4500
Qy	130	-----	129
Db	4501	AVSQSLLSVFQBIILRRGLNGPDPVISTLPLOQGIVDLQROGLLDVQKTEYPRDSSVVDV	4560

QY 130 ----- 129  
Db 4561 PHEQVINPDSIALIHGSEKLSYAQLDRESDRVARLWLRHSFSSDTLIAVLAPRSCETII 4620  
QY 130 ----- 129  
Db 4621 AFLGILKANLAYLPDVKAPAAIDAIVSSLPGNKLLILGANVTPPKLQEAADFPVPIRD 4680  
QY 130 ----- 129  
Db 4681 TFFTLLDGTQLDQPTIERPSAQSLAYAMFTSGSTGRPKGVWVQHNRVRLVKNVNVAKQ 4740  
QY 130 ----- 129  
Db 4741 PAAARIAHISNLAFASSWEIYAPLNGGAIVCADIYFTTIDPQALQETFOBEIRGAMLP 4800  
QY 130 ----- 129  
Db 4801 PSLKQCLVQAPDMISRLDILFAAGDRPSSVDALQAOQLVGSVFNAYGPTENTILSTIY 4860  
QY 130 ----- 129  
Db 4861 NVAENDSPVNGVPIGSAVNSGAYIMDKNQQLVPAGVNGELVVTGDLARGYMPKLDAD 4920  
QY 130 ----- HTAKTITFRP ----- 139  
Db 4921 RPIQLTVNGSEQRAYRTGDRVRPKDFQIEFFGRMDQIKIRGHRIEPAEVEQAFIAND 4980  
QY 140 ----- 139  
Db 4981 GFVEDVAIVIRTPENQPEMVAFTAKGDNARSBEAATQIEGWEAHPGEGGAYANIEIE 5040  
QY 140 ----- PPCGR ----- 144  
Db 5041 SEALGYDFMGWTSYDGTEDKDEMRWLNDRMSLLDGRPAGRVLEVGTCGTMFMNLG 5100  
QY 145 ----- 144  
Db 5101 RSQGLERYIGLEPAPSAAEFVNNAKSPGLAGRAEVHVGTAADVGTQLGLTSDMAVINS 5160  
QY 145 ----- VPSM ----- 148  
Db 5161 VAQYFPTPEYLAETIKSLVQPGMKRIYLGDMRSWAMNDFAAARAAYSADNASKDRVR 5220  
QY 149 ----- TCLSEML ----- 155  
Db 5221 QKMELEKEEELLVDPAFTALASQLQDRIOHQVEIILPKMKATNELSSRYAAVLHISD 5280  
QY 156 ----- NVSKRNDTGE ----- 165  
Db 5281 EPLPIYKIDPEAWINPEGSRLTRELAQVLKENENAESVAISNIPYSKTVVVERHIVRSLD 5340  
QY 166 ----- 165  
Db 5341 QEDANAPEESMDGSDWISAVRTRAQQCHTILSASDLFDIAEDAGFRVEVSWARQHSQHAL 5400  
QY 166 ----- QG ----- 167  
Db 5401 DAVPHLHKPATEDSRVLKPTDQGRPLKSLTNQPLLPQASRRRAELLIREGLQTLPPY 5460  
QY 168 ----- CGNFT ----- 172  
Db 5461 MIPSQITLIDRMPLNANGKVRRELARRAKITQSKPVEDIVPPRNSVEATVCKGFTDVL 5520  
QY 173 ----- TFPNMFNV ----- 181  
Db 5521 GVEVGITDN--FFNLGHSMLATKLAARLGRQLNTRISVRDVPQPVVADLAAVIQNSA 5578  
QY 182 ----- 181  
Db 5579 PHEPIKPADYTPVPQSAQGRMLWFLDQNLNGATWYLMPLGIRLHGLSLRVDALATAISAL 5638

QY 182 ----- 181  
Db 5639 EQRHEPLRTTFPHEBDGVGVQVVDHRRPKDLRIIDLSTQPKDAYLAVLKHEQTTLFLDATE 5698  
QY 182 PRWNTKL----- 188  
Db 5699 PGRVALIRLGEERHILSIWVHHIISDGWSVEVLFDEMHRFYSSALROODPMEQILPLPI 5758  
QY 189 ----- 188  
Db 5759 QYRDPAAWQKTEQVAEBHQRLDYWTEHLADSTPAELLTDLPRPSILSGRANELPLTIEG 5818  
QY 189 ----- 188  
Db 5819 RLHDKLRAFRCRHOATPFVILLAAALRAAHYRLTGAEDATLGTPIANRNRPELENMIGPFV 5878  
QY 189 ----- 188  
Db 5879 NTQCMRTAIBENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQVI 5938  
QY 189 ----- 188  
Db 5939 LAVHSQODLGLKLTLEGLRDEAVDSAISTRPDVEPHLFEHADRLSGSVLYAKELFKLRTIE 5998  
QY 189 ----- 188  
Db 5999 SVVSVFLETLRALDQPLTPLAVLPLTDGVGEIASKGLLDVPRDTPRDANIVEFQOHV 6058  
QY 189 ----- 188  
Db 6059 RATPDAIAVKDATSILTYAQLDQSDRLAIWLSRRHMPETLVGLVAPRSCETIAMPFI 6118  
QY 189 ----- 188  
Db 6119 MKANLAYLPDINSIPAARLSILSAVDGNKLVLLGSGVTAPEQENPEVEAVGIOEILAGT 6178  
QY 189 ----- YV----- 190  
Db 6179 GLDKTQGSNARPSATSLAYVIFTSGSTGPKGVNVHRSVTRLAKPSNVISKLPGARVA 6238  
QY 191 ----- 190  
Db 6239 HLANIAFDASIEIATTLNAGATLVCLDYHTVLDCTRLKEVFERESTVTVTLMALLKQC 6298  
QY 191 ----- GPTKVVNDQSOTIY----- 203  
Db 6299 VAEIPETLAHLDLLTYTGGDRVGGHDAMRARSVKIGMFGSGYGPTENTVIS-TIIEVDADE 6357  
QY 204 ----- 203  
Db 6358 MFVNGVPIGKTVNSGAYVMDRNQQLVPSGVVGVGVVLTGDLARGYTDPSLNKNRFIYIT 6417  
QY 204 ----- 203  
Db 6418 VNGESTRAYTGRVRYRPHDLQIBFFGRMDQVKIRGHRIBPGEVESALLSHNSVQDAA 6477  
QY 204 ----- 203  
Db 6478 VVICAPADQDSGAEMVAFVAARNTEDEDTQEBEAVDQVQGWETHFETAAYSEVKDIRQSE 6537  
QY 204 ----- FLGLTAL----- 210  
Db 6538 VGNDFMGWTSMYDGSIEDKTDMEHLNDTMRMLDAREPGHVLBIGTGTGMFMENLAKCP 6597  
QY 211 ----- 210  
Db 6598 GLQYGVGFEPKSAQAFVNDAAQSPALKDGRSIVHVGTATDINKAGPIQPLRVVINSVA 6657  
QY 211 ----- 210  
Db 6658 QYFPTPEYLPFRVVEALVQIPSVVERIVPGDMRTNAINRDFVNASRALHTLGEKANRQLVRQM 6717  
QY 211 ----- 210

Db	6718	IVELEANE	BELLTDPAFFTSLRGLGEKIKHVEILLPKTMKATNELSKYRYAAVLHVRGSR	6777	Db	7798	AETPELVANLEILHTAGDRDPGDANLAGKTAKGGIFNVLGHTEHTAYSTFYFVVGEBTF	7857
Qy	211	-----	-----	210	Qy	240	ING-----	242
Db	6778	EQSTIHQVSPNAWIDFAADGLDRQTLINLLKHKHDAGTVAIGNIPYSKTIIVERFVNKSLS	6837	Db	7858	VNGVPGRGISNSHAYIIDRHQKLPAGVMGELITGDGVARGVTDTSALNKDRFYVIDIN	7917	
Qy	211	-----	-----	220	Qy	243	-----	242
Db	6838	EDMEEGQNSLDGSAWVAVRVAAGSCPSLDAMDVKEIAQEAQGYQVEVSWARQWSQNGAL	6897	Db	7918	GKSTWSYRTGDKARYRPRDGOLEFFGRMDQVKIRGVRIBEGVELTLLDHKSVAATVW	7977	
Qy	221	-----	-----	220	Qy	243	-----	242
Db	6898	DAIFHHFPPKEGARTLIEFTDYBGRNVNTLTNRPLNSIQSRRLGTQIREKLQTLPPY	6957	Db	7978	VRRPDPGPEMIAFITIDAEDDVQTHKAIYKHLQILPAYMIPSHLVILDQMPVDNGKV	8037	
Qy	221	-----	-----	220	Qy	243	-----	242
Db	6958	MIPSRIMVLDQMPVNNNGKIDRKELVRRRAIVAPKPRSAATRVAPRNEIEAILRDEFEDVL	7017	Db	8038	DRKDALLRAQTVQKRRSTAARVPPRDEBEAVLCEEYSNLLLEVEGITDGFDFLGHSHLLA	8097	
Qy	221	-----	-----	222	Qy	243	TKLK-----NT-----	248
Db	7018	GTEVSVLONFPDLGSHLMATKLAARVSRRLDAHISIKDVDPQVLAADLAASIQRESAPH	7077	Db	8098	TKLAARLSRQLNTRVSKDVDPQIPILADLADIIRRGSHRHDPIPATPYTGPVEQSFAQGR	8157	
Qy	223	-----	-----	222	Qy	249	-----	248
Db	7078	EPIPORPYTGPAEQSFAGRLWFLDQLNLGATWYLMPLAIRGQLRVAALSAAALFALER	7137	Db	8158	LWFLEQLNLGASWYLMPPFAIRMRGPLQTKALAVNALVHRHEALRTTFEDHDGVGVQVI	8217	
Qy	223	-----	-----	222	Qy	249	-----MRKLKXQA-----	257
Db	7138	RHETLRTTFESDGVGVQIVGEARNSDLRVHDVSTGDDGEXYLEVLRREQTVFFDLSEPG	7197	Db	8218	QPKSSQDLRIIDLSDAVDDTAYLAALKRBQTTAFDLTSEPCWRVSLRLGDDDDYILSVM	8277	
Qy	223	-----	-----	222	Qy	258	-----PVKEQ-----	262
Db	7198	WRVCLVKTGBEDHVLISVMHHIYDQMSVDILRGELGQFYSAALRGQDPLHANPLIQY	7257	Db	8278	HHIISDQMTVDVLQELGQFYSAAIRGQEPLSQAKSLPIQYRDPFVWQORQENIKEQAKQ	8337	
Qy	223	-----	-----	222	Qy	263	-----	262
Db	7258	RDFAAWQREAKQVEBHQRLGYSKQLVDSTPAELLTLDRPSILSGRAGSDVDTIEGSV	7317	Db	8338	LKYSQQLADSTPCEFLTDLRPSILSGEADAVPMVIDGTVYQLLTCRTHQVTSFVL	8397	
Qy	223	-----	-----	222	Qy	263	-----	269
Db	7318	YGALQSCFCTRSTVFVLLTVFRIAHFRLTAVDDATIGTPIANRNREPELTIVGCCFVNT	7377	Db	8398	LAAFRTHAYRLTGTLDATVGTPIANRNREPELGLIFFVNTQCMRMAISSETEFESLVQ	8457	
Qy	223	-----	-----	222	Qy	270	TQSTTTPYFS-----	279
Db	7378	QCMRISIADDNFEGVLQVRNVATAAVANQDVPPERIVSALVPGSRNTRSNPLVQLMFA	7437	Db	8458	VLTTTEAFANQDVPPERQIVSTLLPGSRDTSRNPVLQVMFALQSQDGLRIQLEGMTDEA	8517	
Qy	223	-----	-----	226	Qy	280	-----	279
Db	7438	VQSVEDYDQVRLEGLESVMPGEASTRDMFHLVPGDQKLTGSVLYSSDLFEQGTIONF	7497	Db	8518	LETPLSTRLDLEVHLFQEVGKLSGLLYSTDLFEVETIRGIVDVFLERLRRGLEQPKQL	8577	
Qy	227	-----	-----	235	Qy	280	-----	279
Db	7498	VDIFQECLRSVLDQPLTPISVLPSNALSINLESLLLEMPISDYPRDRTVVDLFREQA	7557	Db	8578	MAMPITDGTIKLRDQGLLTIVAKPAYPRESSVIDLFRQVVAAPDAIADWSSSTLTIVADL	8637	
Qy	236	-----	-----	235	Qy	280	-----	279
Db	7558	CPDSIAVKDSSSOLTYAQLDQSDRVAALHERHMPAESLVGVLSRPSRCTIIAYFGIMK	7617	Db	8638	DGQSNKLAHWCORNMAPETLVAVFAPRSCTIVAFGLVKLANLAYLPLDVNAPARIEA	8697	
Qy	236	-----	-----	238	Qy	280	-----	279
Db	7618	ANLAYPLDVYAPDARLAAILDTVEGERLLLLGAGVPQPGIQIPRLSTAVIAEALSHATT	7677	Db	8698	ILSAVPGHKLVLVQAHGPELGLTMADELVIDEALASSSGDHEQIHASGPTATSLAYV	8757	
Qy	239	-----	-----	238	Qy	280	-----	279
Db	7678	VDVTSIPQSATSLAYVIFTSGTGKPGVMIEHGRIVLRVDRTNVNVFPESGALPVSH	7737	Db	8758	MFTSGTGKPGVMIDHRSIIRLVKNSDVVATLPTPYRMANVSNLAFDISVQBIYTALLN	8817	
Qy	239	-----	-----	238	Qy	280	-----	279
Db	7738	FSNLAWDAATWEIYTAVINGTVCIDRDTMLDIAALNSTFRKENVRRAAFTTFAFLKQCL	7797	Db	8818	GGTIVCLDYLTLLDSKILYNVFEAQVNAAMFTPVLLKQCLGNMPAIISRLSVLFNVGDR	8877	
Qy	239	-----	-----	239	Qy	280	-----	279
					Db	8878	LDHDAVAASGLIQDVAVYNAYGPTENGMOSTMWYKVDVNEFFVNGVPIGRSITNSGAYVMD	8937



Qy	280	-----YTTSA-----	284
Db	8938	GNQQLVSPGVMGEIVVTGDLGARGYDTSALDEDFVHVHTIDGEEENIKAYRTGDRVRPK	8997
Qy	285	-----	284
Db	8998	DFEIEFFGMDQOVKIRGHRIEPAVEHALLGDLVHDAVVLKPKANQBEMIAFTSQ	9057
Qy	285	-----	284
Db	9058	EDETIQESHNKQVQGWGEHFDVSRYADI KDLDSTFGHDFLGWTSWYDGVDPVNEKKE	9117
Qy	285	-----	284
Db	9118	WLDETTASLDNRPPGHILEIGAGTGMLSNLKGVDGLQKVGLDPAPSAAFVNEAVKS	9177
Qy	285	-----	284
Db	9178	LPSLAGKARVLVGTALDIGSLDKNEIQPELVVINSVAQYFTSEYLIKVKAVVEVPSVK	9237
Qy	285	-----ALN-----	287
Db	9238	RVFFGDIRSQALNRDFLAARAVRALGDNASKEQIREKIAELESEEEELLVDPAPFVSLSR	9297
Qy	288	-----	287
Db	9298	QLPNIKHVEVLPKLMKATNELSSYRAAVLHISHNEERQLLIQDIDPTAWVDFAATOKDS	9357
Qy	288	-----VTTNVTYSTTAARV-----	303
Db	9358	QGLRNLQOGRDDWMIAGNIPYKTIIVERHIMNSLDQDHVNSLDGTSWISDARSAAAIC	9417
Qy	304	-----	303
Db	9418	TSPDAPALTQAKBEGFRVELSWARQSRONGALDAVFHRLATDANCERSVLVHFPTDQ	9477
Qy	304	-----	303
Db	9478	GRQLTLNRPLQRAQSRRIEBSQVEALQALPAYMIPSRILVLPQMPNTANGKVDKQL	9537
Qy	304	-----	303
Db	9538	ARRAQVAKKAVSARVAPRNDTEIVLCEEYADILGTEVGITDNFDMGGHSLMATKLA	9597
Qy	304	-----	303
Db	9598	LSRRLDTRVTVKEVPDKPVLADLAASIEQGSTPHLPASSVYSGPVEQSYAQGLWFLD	9657
Qy	304	-----	303
Db	9658	QFNLNATWYHMSLAMRLIGPLNMDALDVALRALQRHETLTTPEAQKDIGVQVVEAGM	9717
Qy	304	-----	303
Db	9718	KRLKVLDSLDRKEHMAVLENEQWRPFTLASBPGWKGLARLGTEYILSLVMHHMFS	9777
Qy	304	-----	303
Db	9778	GWSVDILRQELGQFYSAALRGDRPLSQVKPLPIQYRDPFAWQKEAAQVAEHERQLAYWEN	9837
Qy	304	-----	303
Db	9838	QLADSTPCELLTDPRPQFLSGKAGVIVPTIEGPVYEKLLKFSKERQVTLFSLVLLTAPRA	9897
Qy	304	-----STST 307-----	307
Db	9898	THPRLTGAEDATIGTPIANRRNPELEHIIGFFVNTQCMRLLLDTCSTFESLVQHRVSAT	9957
Qy	308	IAY-----	310
Db	9958	DAYSNDQIPFERRIVSALLPGSRDASRSPLIQLMPALHSQDPLGNITILEGHEHRLPTSVA	10017
Qy	311	-----	310
Db	10018	TRFDMEPHLFOEPNKLGSILFADELFPQETINSVTVFQBIILRRGLDQPOVSISTMPLT	10077
Qy	311	-----	310
Db	10078	DGLIDLEKLGILLBIESSNFRDYSVDVPROQVAANPNAPAVVDSETSMSYTSLDQKSEQ	10137
Qy	311	-----RPDS-----	314
Db	10138	IAAMLHAQGLRPESLICVMAPRSPETIVSLFGILKAGYAYLPDVNSPAARIQIPILSEVE	10197
Qy	315	-----	314
Db	10198	GKRLVLLSGIDMPQSDRMDVETARIQDILTNTKVERSDPMRSPATSLAYVIFTSGSTG	10257
Qy	315	-----	314
Db	10258	RPKGWIEHRNILRLVKQSNVTSOLPQDLMAHISNLAFDASIWEIFTAILNGGALICID	10317
Qy	315	-----SPMK-----	318
Db	10318	YFTLLDSQLRRTTEKARVNATLPAPALLKECLNHPATLPEDLVLYIGDRLDATDAK	10377
Qy	319	-----	318
Db	10378	IQALVKGTVYNAYGFTENTVMSTIYRLTQGESYANGVPIGNAVSSGAYIMDKQRLVPP	10437
Qy	319	-----	318
Db	10438	GVMGELVSGDGLARGVTNLTNADRFPVDIVINDQKARAYTGDRTRYRPKDGSIEFFGR	10497
Qy	319	-----	318
Db	10498	MDQVKIRGRHVEPAVEQAMLGKAIHDAAVVQAVDQGETEMIGFVSMASDRFSEEE	10557
Qy	319	-----	318
Db	10558	EITNQVQEWEDHFESTAYAGIEADQATLGRDFTSWTMYNGNLIDKAEMEEWLLDWTQS	10617
Qy	319	-----	318
Db	10618	LLDKEDARPCAEIGTGTGMVLFNLPKNDGLESYVIGEPSRSALFVDKAAQDFPGLQKT	10677
Qy	319	-----	318
Db	10678	QILVGTAEIDIKLVKD PHPD VVVVINSVAQYPPSRSLVQIASELIHMTSVKTIFFGDMRSW	10737
Qy	319	-----	318
Db	10738	ATNRDFLVSRALYTLGDKATKQDQIRQEBVARLEENEDELLVDPAPFTSLTSQWPGVKHVE	10797
Qy	319	-----	318
Db	10798	ILPKRMTSNELSSRYAAVLHICRDGEGRRNRYRRVHSVEENAWIDFASGMRHALVQ	10857
Qy	319	-----	318
Db	10858	MLDERRDAKTVAIGNIPHSNTINERHFTTSLDTEGEGIAQDSLDSAWQSAWKAAARCP	10917
Qy	319	-----	318
Db	10918	CLSVTELVEIQAGFRVEVSUARQSRQHGALDVVPHHLEDDRVGRVLNFTDFERLPP	10977
Qy	319	-----	318
Db	10978	STGLTSRPLORIQRNRPESQIREOLOTLPPYMWPSRIVVLERPLNANSKVDRKELARK	11037
Qy	319	-----SIMATOL-----	325
Db	11038	ARTLOTIKPSATRVAPRNDIEAVLCDEFQAVLGVTVGVGMNDNFELGGHSLMATKLAARLS	11097
Qy	326	-----	325

Db	11098	RLDTRVSVKIDFNQPILODLADVQTGSAPHEAIPSTPYSGPVEQSQRWFLDQLN	11157	Db	12178	VRDVEDTPVVGQAASIOQGSTPHEAIPALSHSGPVQSQFAGRLWFLDRNLNRAWTIM	12237
QY	326	-----	325	QY	357	-----MKNTHVLIR	365
Db	11158	LNASWYHPLASRLRGPLRIEALOSALATIIEARHESLRTTFEEQDGVQIVRAARNKQL	11217	Db	12238	PFQVRLRGPLRVDALQOTALRALERHELLRTTFEEQDGVGMQIVHSPRMRDICVVDISGA	12297
QY	326	-----	325	QY	366	NE-----	367
Db	11218	RIIDVSGTEDAYLAALKQEQDAFDLTAEPGRVALLRLGDDHVLIVMHHIISDGWSV	11277	Db	12298	NEDLAKKEEQAPPNLSTEVAWRVALPKAGENHHILSIIVMHHIISDGWSVDIFQOBLAQ	12357
QY	326	-----ROLATW	331	QY	368	-----	367
Db	11278	DILRQELQOLYNASSOPAPLPIQYRDFAIWQKDSQIAEHQKQNLWKQLVNSKPAEL	11337	Db	12358	FYSVAVRGHDPLSQVKPLPIHYRDFAVWQRQDKQVAVHESQOLWIBQADSTPAEILSD	12417
QY	332	-----VYTTLR	339	QY	368	-----	367
Db	11338	LADTPRKALSGDADVIPIEDDDQVQNLRSFCRHRVTSFVALLAFAAHYRLTGAED	11397	Db	12418	FNRPEVLGSEAGTVPIVIEDEVEYEKLSLFCRHNQVTSFVLLAARVAHVHRLTGAEDATI	12477
QY	340	-----	339	QY	368	-----	367
Db	11398	ATIGSPIANRRPELEGLIGFVNTQCLRIPIVKSEDTFDFLVKQARETATEAQNDQVFF	11457	Db	12478	GTPIANRRPELEDLIGFVNTQCMRIALBEHDFLSVRRVRSTAASAFENQDVPERL	12537
QY	340	-----QNPFF	343	QY	368	-----TPYT	371
Db	11458	ERIVSSWVASSRDTSRNPLVQVMPFAVSHQDLGNIRLEGVEGKPVSMAASTRFDEMHILF	11517	Db	12538	VSALLPGSRDASRNPLVQLMPFVHVSQRNLKQLEGLEGEPTPYTATTRFDFVPHLFEQD	12597
QY	344	-----	343	QY	372	-----	371
Db	11518	EDQMLGNVNVFSKDLFESETIRSVAVFQETLRRGLANPHANLATPLTDGLPSRLSLC	11577	Db	12598	KGLAGNVVFAADLFEAATIRSVVEVFHEILRRGLDQPDIAISTMPLVDGLAALNSRLFA	12657
QY	344	-----	343	QY	372	-----IYGTLD	377
Db	11578	LQVNPDPDASVIDVFREQVASIPKSIADIVASSQLTYTELDERSSQATLWLRQVTV	11637	Db	12658	VEDIEPDPATEASVVDVQTVQVVPANPDALAVTDSTKLTYAELDQOOSHVAAMLKQKLP	12717
QY	344	-----CE	345	QY	378	-----	377
Db	11638	PEELVGLAPRSCETIIAFLGIKANLAYLPLDVNAPAGRIETILSSLPGNRLILLGSDT	11697	Db	12718	AESIVVVLAPRSSETIVACIGILKANLAYLPMDSNVPEARQAILSEIPGEKPVLLGAGV	12777
QY	346	-----	345	QY	378	-----	377
Db	11698	QAVKLHANSVRFRISDALVESGPPTEBELSTRPTAQSLAYVMFTSGTGPVGMVHEHR	11757	Db	12778	PIPDNKTADVVMFISDIVASKTDSYSPGTRPSASSLAYVIFTSGTGRPKGMVHEHR	12837
QY	346	-----	345	QY	378	-----	377
Db	11758	GITRLVNSNVAKOPAAAAIAHLSNIAFDASSWEIYAPLLNGFTVVCIDYTTIDIKAL	11817	Db	12838	VISLVQNASRIPOSLRMAHVSNAFADASVWEIFTLLNGTLFCISYFTVLDLSKALSAA	12897
QY	346	-----	345	QY	378	-----MSSLY	384
Db	11818	EAVFKOHHIRGAMLPALLKQCLVSAPTMISSLEILFAAGDRLLSSQDAILARRAVGSGVY	11877	Db	12898	FSDHRINITLPPALLKQCLADAPSVLSLESLEYIGDRLDGADATKVKDLVKKAYNAY	12957
QY	346	-----	345	QY	385	-----	384
Db	11878	NAYGPTENTVLSTHNGENAFSGVPIGNVAVNSGAFVMDQOQLVSAGIVELVVTG	11937	Db	12958	GPTENSVNSTIYTIETHEFANGVPIGTSGLPKSKAYIMDQOQLVPAGVMGELVAVGDGL	13017
QY	346	-----	345	QY	385	-----E	385
Db	11938	DGLARGYTDKSLVRDFIYITLDGNRVRAVRTGDRVRHRPKDQIEFFGRMDQIKIRGH	11997	Db	13018	ARGYTDPSLNTGRFTHITIDGKQVAVRTGDRVVRPRDYQIEFFGRLDQIKIRGHIE	13077
QY	346	-----	345	QY	386	TMFVENKTASDS	397
Db	11998	RIEPAVEQALARDPAISDAVITOLTDEBEPVAFVFLKGNANGTNGVGVSDQEKID	12057	Db	13078	PAVEQALLSDSSINDAVVSAQNKEGLEVMVGYITTOAASVDKEASNKVQWEAHFDS	13137
QY	346	-----PS	353	QY	398	-----	397
Db	12058	GDEQHALLMENKIRHNLQALLPTWIPSRIIHVDQLPVPNANGKIDRNLAVRAQAPTPTS	12117	Db	13138	TAYANIGGIDRDALGDQDFLSWTSMDGSLIPREEMQEWLNDTWBSLLDNQPGKVLIGT	13197
QY	354	-----SEP	356	QY	398	-----NKTTTSPSM	407
Db	12118	SVSTYVAPRNDIETIICKEFADILSVRGITDNFDFLGHSLLIATKLAARLSRLDTRVS	12177	Db	13198	GTGMVFLNLGVEGLQSVAGLEPSRSVTAMWNKAIEFTFPLAGSARVHVGTAEIDISSIDG	13257
QY	357	-----	356	QY	408	-----GFORTE	413
				Db	13258	LRSDLVWINSVAQYFPPSREYLAELTANLIRLPGVKRIFFGDMRTYATNKDFLVARAVHTL	13317

```
Qy 414 ----- 413
Db 13318 GSNASKAMVROQVAKLEDBEELLVDPAFTSLSDQFPDEIKHVEILPKRMAATNELSSY 13377
Qy 414 ----- 413
Db 13378 RYAAVHVGHQMPNGEBEDKQWAVKOINPKAVDFAGTRMDROALLQLQDRGRDDVV 13437
Qy 414 ----- 413
Db 13438 AVSNIPYSKTIHERHLSQSLDDSDGTSAVDGTAWISRTQRAKECPALSVDADLIEIGK 13497
Qy 414 ----- 413
Db 13498 IGPEVEASWQHSQRGGLDAVHRFBPPRHSGHVMFPTTEHKGSSSLNRLPLHLQ 13557
Qy 414 ----- 413
Db 13558 SRRLEAKVRERLQSLPPYMIPSRITLLDQMLTSNGKVKRKLARQARVIPRSAATLD 13617
Qy 414 ----- 413
Db 13618 FVAPRTIEVVLCSEFTDLLGVKIGITDNPFELGHSLLATKLSARLSRRLDAGITVKQV 13677
Qy 414 ----- 418
Db 13678 FDQPVLAADLAASILQSSRRHSIPSLPYEGPVEQSAQGRLLWFLDQFNIDALWYLIPFAL 13737
Qy 419 ----- 418
Db 13738 RMRGLOVDALAALVALEERHESLRTTFEERDGVGIVQVQPLRTTKDIRIIVSGMRDD 13797
Qy 419 ----- 418
Db 13798 DAYLEPQKEQOTPDPLASEPGWRVALLKLGKDDHILSIVMHHIISDGMSTEVLRBLGQ 13857
Qy 419 ----- 419
Db 13858 FYLAASKGAPLSQVAPLPIQYRDFAVWQREQVAESQRLDYKKQLADSSPAELLAD 13917
Qy 420 Y----- 420
Db 13918 YTRPNVLSGEAGSVSFVINDSVYKSLVFCRSRQVTTFTLLAAFRAAHYRMTGSDDATI 13977
Qy 421 ----- 420
Db 13978 GTPIANRRPELENLIGCFVNTQCMRITIGDDETFESLVQQVRSTTATAFENQDVPFERI 14037
Qy 421 ----- 426
Db 14038 VSTLSAGSRDTSRNPVLQLLFAVHSQQGLGRIQLDGVVDEPVLSTVSTRFLEPHAFQEA 14097
Qy 427 ----- 430
Db 14098 DRLNGSVMFATDLFPETIQGFVAVVEVLQGLEQPSQTIATPLAEGIAQLRDAGALQ 14157
Qy 431 ----- 435
Db 14158 MPKSDYPRNASLVDVFOQQAASPTVAVTSTSKLTYAELDRLSQQAAYLRRQQLPAE 14217
Qy 436 ----- 437
Db 14218 TWAVLAPRSCETIIAPLAILKANLAYMPLDNTPSARMEAIISVPCRRLLIVGSGVRH 14277
Qy 438 ----- 437
Db 14278 ADINVPNAKTMILSDITVTGTDAGTPEPLVVRPSATSLAVIFTSGTGKPKGVMEHRA 14337
Qy 438 ----- 437
Db 14338 IMRLVKDSNVVTHMPPATRMARHVNIAPDVSLEFEMCATLNGTGLVCIDYLLDSTMLR 14397
```

```
Qy 438 ----- 437
Db 14398 ETEFERQVRAAIPFPALLRQCLVNMPDAIGMLEAVYVAGDRFHSRDARATQALAGPRVYN 14457
Qy 438 --PT----- 446
Db 14458 AYGETENAISTYINIDKHPYVNGVPIGSVNSGAYVMDRNOQLLPFGVMGELVVTGE 14517
Qy 447 ----- 446
Db 14518 GVARGYTDASLDTRFVTVTIDGQRQAYRTQDRVRYRPGKFQIEFFGLRDLQQAIRGHR 14577
Qy 447 ----- 448
Db 14578 VELGEVEHALLSENSVTDAAVLRTMBEEDPQLVAFVTTDHEYRSGSNEEDPYATQAA 14637
Qy 449 ----- 459
Db 14638 GDMRKRLRSLPPYMWPSRVITILRQMPNLNANGKVDKDLARRAQMTPTASSSGPVHVAPR 14697
Qy 460 ----- 459
Db 14698 NETEAAICDEFETILGVKVGITDNFPELGHSLLATKLAARLSRRMGLRISVKOLFDDPV 14757
Qy 460 ----- 459
Db 14758 PVSILAGKLEQQQSGEDESSTVGIVPQLLPAEMSREIIQDVVPOIENGHSTPLDMYP 14817
Qy 460 ----- 459
Db 14818 ATQTIFFLHDKATGHATPPLSLDFPETADCRRLASACALVQHFDIFRTVFSRGR 14877
Qy 460 ----- 459
Db 14878 FYQVVLHLVDVPEVETEQELDEVALALHEADKQPLRLGRAMLRIAILKRPCKAKRLV 14937
Qy 460 ----- 461
Db 14938 LRMSHSLYDGLSLEHIVNALHALYSKHLQAQPKPGLYMHMMSRRABGYNFRSILQGS 14997
Qy 462 ----- 462
Db 14998 SMTSLKRSVGALEAMTPSAGTWTQTSKIRIPPAALKNGITQATLPTAAVSLLLAKHTKST 15057
Qy 463 ----- 462
Db 15058 DVVFGVRVSGRQLSINQDIVGPCINEVPVRVIDEGDDMGLLRAIQDQVTSFRHET 15117
Qy 463 -----W 463
Db 15118 LGLQEVKENCTDW 15130
```

## RESULT 4

T30192  
probable peptidase synthetase - Aureobasidium pullulans  
C:Species: Aureobasidium pullulans  
C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: T30192  
R:Peery, R.B.; Thornnewell, S.J.; Tobin, M.B.; Skatrud, P.L.  
submitted to the EMBL Data Library, January 1997  
A:Description: Discovery of an MDR-like gene adjacent to a peptidase synthetase in Aureoba  
A:Reference number: Z20767  
A:Accession: T30192  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-10797 <PEB>  
A:Cross-references: UNIPROT:O94116; EMBL:U85909; NID:G4099310; PID:G4099313; PIDN:AA0005  
C:Genetics:  
C:Keywords: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2  
C:Keywords: carrier protein  
F:1618-1688/Domain: acyl carrier protein homology <ACPI>  
F:3682-3752/Domain: acyl carrier protein homology <ACP2>

F;5615-5685/Domain: acyl carrier protein homology <ACP3>  
F;7503-7573/Domain: acyl carrier protein homology <ACP4>  
F;9683-9752/Domain: acyl carrier protein homology <ACP5>

Query Match 22.6%; Score 555; DB 2; Length 10797;  
Best Local Similarity 2.2%; Pred. No. 0.047;  
Matches 227; Conservative 105; Mismatches 127; Indels 9990; Gaps 73;

QY	1	MGRKEM-----	6
Db	84	ISRKDLGIVPIYIPQCISLYISQALILKGAAPCLNLDVPGERLKFILKOTSASILLTT	143
QY	7	-----	6
Db	144	AAMRPKLELEGITVIIVVDELSEFKSDREPAGEKITSPDTSSLAYIMTSGTGIPKAV	203
QY	7	-----	6
Db	204	CLSHRAVTSQSLAHDRFIPFSRFLQFASPTFDVSVEIFFPWPYRGATLVSVERNLLGD	263
QY	7	-----MVR---DVPKMFVLIS-----	19
Db	264	LPFTTSLNIDAAELTPSVAASLVHRHENVPTLRLLTTIGEMLNTQVIQFGSGIKSGI	323
QY	20	-----	19
Db	324	LYGMVGPTEAIIHCTLQPGVDLPAGTIGIPLDTVSCFIVKPTSTKHASQLEILPIGE	383
QY	20	-----	19
Db	384	IGELVIGHQLAGVYLNREEOTRAAFVTHPKFGGLYRTGDKARLHRNGTLECYGRISGQ	443
QY	20	-----ISFLVSPINC-----	30
Db	444	VKLGRQVRELGEIEHAASKAGCHAVIASVISGLVLP--CIGDHRVSSKDIKSAOCWK	501
QY	31	-----KVMKAL---YN-----	39
Db	502	LPAYMIPSDIVLLDDFPYLPKGVKDKRLTSDYNSNTAQHVSGSDLSENAREITRIES	561
QY	40	-----	39
Db	562	VLGVSDHSDTLASAAGLSLRAIQVASQLRRQGCADLGALELLSVSNVLADELVRKAD	621
QY	40	-----RP---	41
Db	622	ESNINDNDSEKWKQTVHELRSVERDPESKAFVSGIEDVLPTPLQDAMLVETAKRPOAY	681
QY	42	-----WRGLVLSKI--	50
Db	682	CNELRTVSPKIPVERVRQALFALAQHRTALRSGFMPSGVSHCAVTVQVIMKTLVTSQFAH	741
QY	51	-----	50
Db	742	VKSFTTGWSVTNRETLRLPLHFQYKCSGAEBILVAIHIALYDQMSVEVILEDETLLQN	801
QY	51	-----GKYKLD-----	56
Db	802	ERTPERPSGAVNKFFNLRRSEDTQSHLDFWGEVLSVTPGRLNLSFGKMPQPQLOSIQ	861
QY	57	---QLKLEILRQ-----	65
Db	862	HTIEMDETLRQAASHYSCSAHVFFQAYAILLGFYMGTEDTVFGTVFSGRTLPIVIES	921
QY	66	---LETTISTKYN-----	75
Db	922	MVGPLLSTLPTPRINTLESRKPSDVLRLQEDNRKIMRHSMTSLADIKKACGFNPGEAVFD	981
QY	76	-----	75
Db	982	SIFVQETARPDAQAQTLLNLVEAHDYLEFNLTLELPTQQGVKTKATYQSSLLPLQHVK	1041
QY	76	-----VSKQ-PVKQLTWNTPEPQYIIL-----	96

Db	1042	TLLQQLDALVKIVVARPETHMNEISDQLPISLVSEANSEPOSFVYKAGLSLVENHALNN	1101
QY	97	-----AGPIQNYGIT-----	106
Db	1102	SGGLALVFAHDIREGTSRMESLTYGELNTRANQLANLYLISQAKRDELICVMEKSVSLY	1161
QY	107	-----	106
Db	1162	LSILAAVKAAGCYLPLVPETPAARIQILAEADVKFCCLTDSMAPVIADVSRCHIMNVD	1221
QY	107	-----YLMF-----	106
Db	1222	TDCSAQSGTGPQLDFKPTDIAIYAVFTSGTTGKPKGVLVTQENILSNLEVLISKIYVPEGS	1281
QY	107	-----YLMF-----	110
Db	1282	RLAQACQAPDVSVFEIFFTWYTCMLCSASKDVMFRDPEKAINELITHLSLTPVAAL	1341
QY	111	-----	110
Db	1342	TDPAHIPRVKFLVTAGEAVTHHVHAGWAGKLYQGYGPSETTNICTVNSAVESDHVINNI	1401
QY	111	-----	110
Db	1402	GPAFENTSAPVLTOGDDFQPLVPLGGELGELCFGGQQVFRGYQNMPELTESKIINHNYGRI	1461
QY	111	-----	110
Db	1462	YRSGDLGRLLPDGTILIQGRTDQKIRGQRIELGEISGCLLPSPVQNCALIEVIKTADK	1521
QY	111	-----DFYS-----	114
Db	1522	ERLMAFWIPSGYSKDSYILQPKNLEBIIKSIYAHADNLPAVMVPDALVPVSAIPQTS	1581
QY	115	-----	114
Db	1582	QKIDKXERLASDGSALTVEDLNAYSGADDDTSELSATEQQALASALADTLQMSQTSIGR	1641
QY	115	-----TQIRKPKAY-----	123
Db	1642	STSFALGLDSVAIRLATNLRLKEYGYSIDVSQILKRPTIARLAPLGGESSKQNEPVT	1701
QY	124	-----VVSQVNTAKTIT-----	136
Db	1702	ADCEAAVGSYLHDESVVSQLHEHGQTVSQVLPTPLQEAMLSARDTSGSSAYRNKTLFSL	1761
QY	137	-----	136
Db	1762	HGSVDKILKACWEVMLQRHDLRTIFLSTEDSRFPFVQAVLSQWTLPWQECDDIPDQLSTL	1821
QY	137	-----	136
Db	1822	LDSAKAGDSIVDHPSPWKIQVYRSESTVYLLDMHHALYDANAMSNLLYEVEQLYKQDS	1881
QY	137	-----PRPPP-----	141
Db	1882	LSAPVSFKPLNPMISTSVBEADALFRDQLREFVPKPRKTRTDVKSGFGTITGRNLNYSKM	1941
QY	142	-----CGRV-----	145
Db	1942	VETFLSKHSTTMLSLTOAMMMKTLAASQSVSDVCCGVNVSGRSVPVDGIESLVAPCFNII	2001
QY	146	-----	145
Db	2002	PVRVLSKHSRNLGLVKALQVRNIDSLPYQLTPLRRIOAQAGTNKRLFDLSLVLLQDDTT	2061
QY	146	-----	145
Db	2062	DLDSAIRLEGESGVMDDETILNPLTLLEGPELLEHISGTPETSCTALEYLDADOK	2121
QY	146	-----	145

Db 2122 TITSLYDLHRRSEHLAWRLIEVSRHSKDLGIPIYIPOCISLYISQLAILKSGAFC 2181  
QY 146 ----- 145  
Db 2182 PLNLDPGERLKFILKDTASILLTTAAMRPKLPELEGITVIVDDELSEFKSDREPAGE 2241  
QY 146 -----PSMTCLS----- 152  
Db 2242 KITSPDSSLAYIMYTSSTGLPKAVCLSHRAVTOSSLAHDRFIPFSRFLQFASPTPDV 2301  
QY 153 ----- 152  
Db 2302 SVPEIFPPMYRGATLVSVERNRLGLDLPGTITSLNIDAEALTPSVAASLVRHRENVPTLR 2361  
QY 153 -----EMLN----- 156  
Db 2362 ALLATIGEMLNTQVIQFGSGIKSGILYMGYPTEAAIHCTLQPGFVDLPAGTIGIPLD 2421  
QY 157 ----- 156  
Db 2422 TVSCFIVKPTSTKASQLEILPIGEIGELVIGGHQADGYLNREBQTRAFAFVTHPKFEG 2481  
QY 157 ----- 156  
Db 2482 LYRTCDKARLHRNGTLECYGRISGQVKLRGORVELGEIEHAASKAGCHAVIASVIGL 2541  
QY 157 ----- 156  
Db 2542 LVLFICGPHRVSSKDIKSAQKMPAYMPIPSDIVLLDDFPVLPBGKVDKRLTDYNSN 2601  
QY 157 ---VSKENDTGE-----QCGG- 169  
Db 2602 TAQHVSGSSDISENAREITRIESVLGVSIDHSTDLSAAGLDSLRAIQVASQLRRQGCAD 2661  
QY 170 ----- 169  
Db 2662 LGALELLSVSNVALDELVRKADESINNDSEKWKQTVHELRSSEVERDFESKAFVSGI 2721  
QY 170 ----- 169  
Db 2722 EDVLPCTPLQDAMLVETAKRPOAYCNEURLTVSPKI PVERVROALFALAQHRTALRSFGM 2781  
QY 170 ---NETT---FNM----- 177  
Db 2782 PSGVSHCAYQVIMKTLVTSQFAHVKSPTTGMVSNRETLLRPLHFQYKCSGABAEILVA 2841  
QY 178 -----FENVPR-----WNTKL- 188  
Db 2842 IHHALYDQMSVEVILEDLLETLLQNERTPERPSPFGAVNKFPLNRRSEBQTSHLDFWGEYLS 2901  
QY 189 -----Y 189  
Db 2902 DVTGRLPNLSPKMPPOPLQSIQHTIEMDMETLROAAHSYSCSAHVFFQAAVAILLGFY 2961  
QY 190 VG-----PTKNV----- 197  
Db 2962 MGTEDTVFTVSGRTLPIVEIESMVGPLSTLPTRLNLESRKFSFVLSRLQEDNRKIM 3021  
QY 198 ---DS-----QT----- 201  
Db 3022 RHSMTSLADIKACGFNPGEAVFDSIFVMQETARPDARAQTLNLNVEAHDYLEFNLTLEL 3081  
QY 202 ----- 201  
Db 3082 EPTQGVKTKATQSSLLPLQHVKTLLQDLALVKIVVAREPETHNNEISDOLPISVLSVA 3141  
QY 202 ---IYFLGLTALLRYAORN----- 218  
Db 3142 NSEPOSFVYKAGLGLSVENHALNNSGGLVFAHDIREGTSMESLTATVGEINTRANQLAN 3201  
QY 219 ----- 218  
Db 3202 YLISQGAKRDELI CVMKEKSVLSYLSTILAAVYKAGCGYLPLVPETPAARIQILAEADVKF 3261

QY 219 ----- 218  
Db 3262 CLTDSSMAPVIADVSRCHIMNVDTTDCSAQSGCTGQLDFKPTDIAYAVFTSGTKPKGV 3321  
QY 219 ----- 218  
Db 3322 LVTQENILSNLEVLISKYIPVPEGSRLLOACNQAFDVSVEIFEPTTWTGMCCLCSAKSADVMF 3381  
QY 219 ----- 218  
Db 3382 RDEKAINELLEITHLSLTPPTVAULTDPAH IPRVKFLVTAGEAVTHHVHAGWAGKLYQGY 3441  
QY 219 ----- 218  
Db 3442 GPSETTNICTVNSAVESDHVINNIGPAFENTSAFVLTOGDDFQLVPLGLGELCFGGQOV 3501  
QY 219 ----- 218  
Db 3502 FRGYONMPELTESKI INHPNYGRIYRSGDLGRLLPDGTILIOGRTDDQKIRGQRIELGE 3561  
QY 219 ----- 218  
Db 3562 ISGCLLQPPSVONCAIEVIKTADKERLMAFWIPSGYSKDSYSLIQPDKNLEELIKSIYAH 3621  
QY 219 ----- 218  
Db 3622 LADNLPAYMPDALVPVSAIPQTSQKIDKRRLASDGSALTVEDLNAYRGADDDTSEL 3681  
QY 219 ----- 218  
Db 3682 SATEQQLASALADTLMQSQTSIGRSTSFALGLDSVAIRLATNLNRKEYGYSIVDSQILK 3741  
QY 219 ---CT--- 220  
Db 3742 RPTIARLAPLLGESSKQNEPVTADCEAAVGSYLHDESVVSQLHEHQTVSQVLPCTPL 3801  
QY 221 ----- 220  
Db 3802 QEAMLSARDTSGSAYRNKTLFSLHGSVDKLKACMEVLMQRHDLRLTIFLSTEDSRFPFV 3861  
QY 221 ----- 220  
Db 3862 QAVLSOWTLRQWECDDIPDQLSTLLDSAKAGDSIVDHPWPKIQVYRSESTVYLLDMH 3921  
QY 221 HSPFLVNMGRNLPV----- 236  
Db 3922 HALYDANMNSNLLYEVEQLYKQSLAPVSFKPFLNFMISTSVEEADALFRDQLREFVPK 3981  
QY 237 ---PKYI--- 240  
Db 3982 PKRTDVKSGFTITGRNLNYSKMWETFLSKHSTTMLSQTQAMMKTLAASQSYSDVCCG 4041  
QY 241 ----- 240  
Db 4042 NVWGRSVVPDGIESLVAPCFNTIPVRVDLSKHSRNLGLVKALQVRNIDSLPYQLTLRR 4101  
QY 241 ----- 240  
Db 4102 IQAQAGTNGKRLFDLSVLLOQDPTDLSAIWRLEGESGVMDMRPKLPELEGITVIVUDE 4161  
QY 241 ----- 240  
Db 4162 LSEFKSDREPAGEKITSPDTSLLAYIMYTSSTGLPKAVCLSHRAVTOSSLAHDRFIPSF 4221  
QY 241 ----- 240  
Db 4222 SRFLQFASPTFDVSVEIFEIPFPYRGATLVSVERNRLGLDLPGTITSLNIDAEALTPSVA 4281  
QY 241 ----- 240  
Db 4282 SLVRHRENVPTLRALLTIGEMLNTQVIOQFGSGIKSGILYMGYPTEAAIHCTLQPGFG 4341

QY 241 ----- 240  
Db 4342 VDLPGTIGIPLDTVSCFIVKPTSTKHASQLEILPIGEIBELVIGGHQADGYLNREEQ 4401  
QY 241 -----NGT----- 243  
Db 4402 TRAAVTHPKFGLVYRTGDKARLHRNGTLBCYGRISGQVKGQRVBLGHEHAASKAG 4461  
QY 244 ----- 243  
Db 4462 GCHAVIASVISGLLVLCIGDPRHVSSKDIKSAQKWLPAWIPSDIVLLDDPPYLPSCG 4521  
QY 244 ----- 243  
Db 4522 VDKRLETDYNSNTAQHVGSSDLSNAREITRIIESVLGVSIDHSTDLSAAGLDSLRAI 4581  
QY 244 -----KLMKMKLKKQ 256  
Db 4582 QVASQLRRQGCADLGALELLSVSNVLADELVRKAKADESNINDNSEKWKQTVHEL- 4638  
QY 257 APVKEQPEKK-----AKTQS-----TTTP----- 276  
Db 4639 SSVRFESKAFVSGIEDVLPCTPLQDAMLVETAKRPOAYCNELRLTVSPKIPVERVROA 4698  
QY 277 ----- 276  
Db 4699 LFALAQRHTALRSGFMPGSHCAVTVQVWKTIVTSQFAHVKSFTTGWSVTNRRETLRLPL 4758  
QY 277 YFSYTTSA----- 285  
Db 4759 HFQYKCSGAEAILVAIHALLYDQMSVEVILEDTLLQNERTPERPSGAVNKFNLRR 4818  
QY 286 ----- 285  
Db 4819 SEDQSHLDPMGEYLSVDTVPGRLPNLSPKMPPQPLQSIQHTIEMDMETLROAAHSYCS 4878  
QY 286 ----- 285  
Db 4879 AHVFFQAYAILGPFMGTEDTVFGTVPFSGRTLPIVEIESVGVPLLSLTPRINTLESRK 4938  
QY 286 ----- 285  
Db 4939 FSDVLSRLQEDNRKIMRHSMTSLADIKKACGFNPEAVFDSIFVMQETARPBARAQTLLN 4998  
QY 286 -----LNV----- 285  
Db 4999 LVEAHDYLEFNLTELEPTQGVTKATYQSSLLPLQHVKTLLQQLDALVKIIVARPETH 5058  
QY 286 ----- 285  
Db 5059 MNEISDQLPISVLSVANSEPOSFVYKAGLSLVENHALNNSGGLALVPAHDIREGTSRME 5118  
QY 286 ----- 285  
Db 5119 SLTYGELNTRANQLANYLISQAKRDELICVMEKSVSLYLSILAAKAGCGYLPVPET 5178  
QY 286 -----LNV----- 288  
Db 5179 PAARIQILAEADVFKCLTDSMAPVIAVDSRCHIMNVDTTDCSAQSGTGPQLDFKPTDI 5238  
QY 289 ----- 288  
Db 5239 AYAVFTSGTGKPGVLVTQENILSNLEVLISKIYVPVPGSRLLQACNQAFDVSVEIFFT 5298  
QY 289 ----- 288  
Db 5299 WYTGMLCSASKVMFRDFEKAINELITHLSLPTVAALTDPAPHPRVKFLVAGEAVT 5358  
QY 289 -----TTNV----- 292  
Db 5359 HHVGAWAGKGLYQGYGSPSETINICTVNSAVESDHVINNIGPAPENTSFAVLTOGDDFQL 5418  
QY 293 ----- 292

Db 5419 VPLGGLBELCFGQOQVFRGYQNMPELTESKIINHPNYGRIVRSGDLGRLLPDGTILIOQR 5478  
QY 293 -----TYSI- 296  
Db 5479 TDDORKIRGORIELGEISGCLLQFPFVQONCAIEVIKTADKERLMAFWIPSGYKSDSYIL 5538  
QY 297 ----- 296  
Db 5539 QPDKNLBEIKSIYAHILADNLPAWVPDALVPVSAIPQTSQKIDKRRLASDGSALTYVED 5598  
QY 297 -----TTAARRVSTS 306  
Db 5599 LNAYSRGADDDTSELSATEQOLASALADTLQMSQTSIGRSTSFALGLDSVAIRLATN 5658  
QY 307 -----TIA----- 309  
Db 5659 LRKEYGYSIDVSOILKRPTIARLAPLLGGESSKOTNEPVTADCEAAVGSYHLHDSVWSQL 5718  
QY 310 ----- 309  
Db 5719 HEHGQTVSQVLPCTPLQENMLSARDTSGSSAYRNKTLFSLHGSVDKLCACWEVMLQRHDI 5778  
QY 310 -----YRPDSS----- 315  
Db 5839 QVYRSESTVYLLDMHHALYDANAMNLLYEVOQYKQSLSAPVSPKPLNFMISTSV 5898  
QY 316 -----F 316  
Db 5899 EADALPRDQLREFVPKPKRTDVKSFGTITGRNLNYSFKQVETPLSKHSTMLSLTQAMW 5958  
QY 317 MKSIMATQ----- 324  
Db 5959 MKTLAASQSYSDVCCGNWVSGRSVPDGIESLVPFCFNTIPVRVDLSKHSNGLVKALQ 6018  
QY 325 -----LR-----DLATW-----VYT 334  
Db 6019 RVNIDSLPYQLTPLRRTIOAQAGTNGKRLFDSLVLLOQDQTTDLDSAIWRLEGESGVMDMYT 6078  
QY 335 T----- 335  
Db 6079 SGTGLPKAVCLSHRAVTQSLAHDRTFIPSPSRFLQFASPTFDVSVFEIPFPWYRGATLV 6138  
QY 336 -----LRYRNP----- 342  
Db 6139 SVERNLLGDLPGTITSLNIDAAELTPSVAASLVHRHENVPTLRALLTIGEMLNTQVIOQ 6198  
QY 343 ----- 342  
Db 6199 FGSGIGSGILYGMGYPTAAIHCTLOPGFVDLPAGTIGIPLDTVSCFIVKPTSTKHA 6258  
QY 343 ----- 342  
Db 6259 SOLEILPIGEIBELVIGGHQADGYLNREBQTRAAVTHPKFGLYRTGDKARLHRNGTL 6318  
QY 343 -----FC----- 344  
Db 6319 ECYGRISGQVKLRQORVELGEIEHAASKAGGCHAVIASVISGLLVLCIGDPRHVSSKD 6378  
QY 345 ----- 344  
Db 6379 IKSACQKWLPAWIPSDIVLLDDPPYLPKGVKKRLETYNSNTAHOVSGSSDLSNAR 6438  
QY 345 ----- 344  
Db 6439 EITRIIESVLGVSIDHSTDLSAAGLDSLRAIQVASQLRQGCADLGALELLSVSNVLALD 6498  
QY 345 ----- 344

Db 6499 ELVRAKADSNINDSEKWKOTVHELRSVERDFESKAFVSGIEDVLPCTPLQDAMLVE 6558  
Qy 345 ----- 344  
Db 6559 TAKRPOAYCNELRLTVPKIPVERVRQALFALAQRHTALRSFGMPGSHCAYTQVIWKT 6618  
Qy 345 ----- 344  
Db 6619 LVTSQFAHVKSFTTGWSVTNRRETLRLPHFQYKCSGABAILVAIHHALYDQMSVEVILE 6678  
Qy 345 ----- 344  
Db 6679 DLETLQNERTPERSFGVANKFPNLRSEDOTSHLDFWGEVLSVDTFGRLPNLSPKMP 6738  
Qy 345 ----- 344  
Db 6739 POPLQSIQHTIEMDMETLQAAHSYCSARVFFQAAVAILLGFYMGTEDTVFGTVFSGR 6798  
Qy 345 ----- 344  
Db 6799 LPIVEBSMVGLLSTLPTRLNLTLESRKFSVLSRLQBDNRKIMRHSMTSLADIKKACGF 6858  
Qy 345 ----- 348  
Db 6859 NPGEAVIDSI FWOETARPDARAQTLNLNVEAHYLEFNLFILELEPTQOGVTKATYQSS 6918  
Qy 349 ----- 348  
Db 6919 LLPLQHVKTLLQOLDALVKIVVARPETHMNEISDQLPISVLSVANSEBPQSFYKAGLSL 6978  
Qy 349 ----- 348  
Db 6979 VENHALNNSGGLALVFAHDIREGSRMESLTYGELNTRANQLANYLSOGAKRDELICVC 7038  
Qy 349 ----- 348  
Db 7039 MEKSVLSYLSILAAVACGGYLPVPETPAARIQILAEADVCFCLTSSMAPVIADVSR 7098  
Qy 349 ----- 348  
Db 7099 CHINNVDTDCSAQSGCTGPDLPDIAYAVFTSGTGKPKGVLVTOENILSNLEVLK 7158  
Qy 349 ----- 348  
Db 7159 IYPVEGSRLLQACNQAFDVSVEIFFTWYTMCLCSAKDVMFRDFEKAINELEITHLS 7218  
Qy 349 ----- 358  
Db 7219 LPTVAALTDPAHPRVKPLVTAGEAVTHVHGWAGKLYOGYGPSETTICTVNSAVE 7278  
Qy 359 NTHVL----- 363  
Db 7279 SDHVINNIGPAPENTSAPVLTOGDDFQLVPLGGLGELCFGQGVFRGYNMPELTESKII 7338  
Qy 364 ----- 363  
Db 7339 NHPNYGRIYRSGDLRLLPDGTILIQGRDQDKIRGRIELGBISGCLLOFPSPVONCAI 7398  
Qy 364 ----- 363  
Db 7399 EVIKTADKERLMAFWIPSGYSKDSYSILOPKNLEBIKSYAHADNLPAYMVPDALVP 7458  
Qy 364 ----- 363  
Db 7459 VSAIPQTSQKIDKRRLASDGSALTVEDLNAYSRGADDDTSELSATEQQLASALADTLQ 7518  
Qy 364 ----- 363  
Db 7519 MSQTSIGRSTFFALGLDSVAIRLATNLKEYGYSIDVSIQILKPTIARLAPLLGESS 7578  
Qy 364 ----- 364  
Db 7579 KOTNEPVTADCEAAVGYLHDESVVSQLHERGQTVSQLPCTPLQEAMLSARDTSGSSAY 7638

Qy 365 RNETPYTYITGTD----- 377  
Db 7639 RNKTLFSLHSGVDLKLKACWEVMLQRHDLRTIFLSTEDSRFPFVQAVLSQWTLPWQECDD 7698  
Qy 378 ----- 381  
Db 7699 IPDQLSTLLDSAKAGGSIVDHSPPKIQVYRSESTVYLLDMHHALYDANAMSNLLYEV 7758  
Qy 382 ----- 381  
Db 7759 EQLYKQSLASAPVSPKPLNFMISTSVEEADALFRDQLREFVPKPFKRTDVKSFGTITG 7818  
Qy 382 ----- 381  
Db 7819 RLNYSPKMVETFLSKHSTTMLSITQAMMKTAAASQSYSDVCCGNVSGRSVPVDGIESL 7878  
Qy 382 ----- 381  
Db 7879 VAPCNTIPVRVDSLKHRSNLGLVKALQRVNIDSLPYQLTPLRRIQAQAGTNGKRLFDL 7938  
Qy 382 ----- 381  
Db 7939 VLLQDQTTDLDSAIWRLEGESVMDPCIVELAPTNESYTLSLHFNRSYLDDEVVSNLHQ 7998  
Qy 382 ----- 381  
Db 7999 ACLSAFASCIRYPSDSVDFIDFDADLVAGVLKPTKHMQPVEAAKTNRSEKSGSGDES 8058  
Qy 382 ----- 381  
Db 8059 WSPLELOIRAAYSAVSSAPEDRIRDDTTYIKGLDLSIAIQLANRLRKDGLLVQASDME 8118  
Qy 382 ----- 381  
Db 8119 SPSCSELASAVQSSQTPVLDERGDFEGDPKHYRGAALQSHRIATEKVASVRPCTPLQS 8178  
Qy 382 ----- 388  
Db 8179 GMLSEYTHSDGHQYFNHTFYAIEADIDSSKLQSAWSKVLEQHELLRTGTVGTDDEHPFV 8238  
Qy 389 ----- 388  
Db 8239 MLTYTFDVIDLEIQASSREGSVYSEKQKASESVQNNHLPPMRWSLLGVEGRQCLQFS 8298  
Qy 389 ----- 388  
Db 8299 AHHAIPDAESLRLIMTDLQSALSNGYVPTRLTIDGALGHILSNQADVESORTFWSQKLS 8358  
Qy 389 ----- 388  
Db 8359 GAPVTRFPNMTFVRISDTEAANVELVLYKRSKLEARCOELGVSMQSVGQAARLLSAY 8418  
Qy 389 ----- 396  
Db 8419 TGESQVTFGVVLSGRTSPATADAPPCITTLFVSTNTAVDDSQFLKOLMSYNATIQKHQF 8478  
Qy 397 ----- 396  
Db 8479 TPLTNIRYABSTSEALFDSLFTVYQRPNNVDVSSWKIIREKASVELAVSMEALSED 8538  
Qy 397 ----- 396  
Db 8539 GLGLRLTIDPAQVPHEQKIMLQOMEVMIAGLLAKFADAINTSVMIIIPPKPIIATDFKY 8598  
Qy 397 ----- 401  
Db 8599 LHENTEASVKSVDRIAMEFVDALEDGQIISRHWTYRQDDEANKIAHLLIDRGVKGDI 8658  
Qy 402 ----- 406  
Db 8659 IATSFDKCPASPAFYGILKAGCAFCADPTAPARKAFILEDNARVLLTSDIRSRL 8718

QY 407 ----- 406  
Db 8719 ELTQCDIIDLINFENKNELTSSVPVSGLAPSSVSYLYTSGTGTGPKGCEITHDNAVL 8778  
QY 407 -MGQRTF-  
| | | |  
Db 8779 VMSFKRLFKGRWTDSEWLQFASVYHFDVSVLEQFWTWVGMRLVCAPRDLILEDIAGFLD 8838  
QY 414 -----TDPL----- 417  
| | | |  
Db 8839 TMQITHLDLTFSLGRLLDPALVPSLHKGVFITGESLKQDQINTWGDVGLCFNFGFTFC 8898  
QY 418 -----WD----- 419  
| | | |  
Db 8899 TIGVTVPFVCKEGKPSNIGMQFDNVGCVYVLAGTQTPVLRGAIGELCISGKLVKGYNL 8958  
QY 420 -----YLSL-----LFLDEIR----- 431  
| | | | | | | |  
Db 8959 RPELTADCFPYLDAGFGRVYRTGDLVRLFHDGSDIFLGRKDNQVQLRGQRLEIDEIAVI 9018  
QY 432 ----- 431  
| | | |  
Db 9019 KRCQDIQDTCIVAKHPKQDKQOLIAFIGINESRKQKPELCPAESTRHLIQTARAACEE 9078  
QY 432 ----- 431  
| | | |  
Db 9079 RLFYGMVPTHFLPIQRIPLSVNNKVEKLLRLQYADLPTTVIQTATQADSQSLSDGEQ 9138  
QY 432 -----NFS-----LRSPTVV 441  
| | | | | | | |  
Db 9139 KVAQALABELLKIDDNDLTSSNIFSLGLSSIQAIFSKKLKASGFTTVQATVLRKNTIS 9198  
QY 442 NLT----- 444  
| | | |  
Db 9199 RLTKALATSTGRSGGIBADAKQVISACRQRMGTVTRVLRCCKADDIEAIPCTPLQOQII 9258  
QY 445 ----- 444  
| | | |  
Db 9259 SRSLASESLYFNSFKFNAQGVLDLQKLGAFNAQLERTQILRTFPIETDDGVQAVRKTG 9318  
QY 445 ----- 444  
| | | |  
Db 9319 HLPWMTLEVYDLASVDGVPAKQKQWRSYNTSHLTVPEIVIVRSGETFVSVDLHLY 9378  
QY 445 -----PP 446  
| | | |  
Db 9379 DGNSTFILLNNVSKLYNSQADFGKPFVDCLAFGLPLNVQAKQFWLHLPDVKSASMP 9438  
QY 447 ----- 446  
| | | |  
Db 9439 LIDNPASHDVLCTASLDILNQADELRSLGVTVOALVOATWATLKRHYQGAIGTVVSGR 9498  
QY 447 ----- 446  
| | | |  
Db 9499 SIDFDGVENVIGLENTIPFYLRCPEGTWQLTVQRCHDFNTALPYQHTPLRDIVKCN 9558  
QY 447 ----- 446  
| | | |  
Db 9559 KHSRPLDALFVYQGTLDNSDTNHSILKPLEDDSFADYPLSPAEAEAGGNLKISVAA 9618  
QY 447 ----- 446  
| | | |  
Db 9619 KASICNETKARELIDEFHQAFAMNKSPENEVGASIGHFPERLARTEDGTKEIQARDTS 9678  
QY 447 ----- 446  
| | | |  
Db 9679 DFHWGSEASVIRSEIATLAGVQEQADIDERTSIFEVGLDSVDVAVKLSRLKKGVLPVST 9738  
QY 447 ----- 446  
| | | |  
Db 9739 IMRSQTIAQWVLTITNGKTDTSKAKSGAFALESKLTSTVARDQINGDKSFERVLPSPM 9798  
QY 447 ----- 446  
| | | |

Db 9799 QBALVSEMVRTSYKAYFNHVDVLKLSKELDIVRLEQAWRTVIKVSPILRTGLEIDDPDLN 9858  
QY 447 -----EHRRAVNLSTN----- 458  
| | | | | | | |  
Db 9859 ATFAQIIQKAGFADPDHKLKSSDAINEHLAHINSVSSAALFQPLRLTRVSTPEDNYL 9918  
QY 459 ----- 458  
| | | |  
Db 9919 VLSVAHALYDGHSLSLHLDVQNAVNETFQSRPFVADILRDTFHGSNEESVWFWRSLTD 9978  
QY 459 ----- 458  
| | | |  
Db 9979 AKKTQLPOIKNEGSVTTTHRTKTSRLAANDLSSFCQKQGVTLQAACQTAWAFVAHTV 10038  
QY 459 ----- 458  
| | | |  
Db 10039 QSLDVMGVVLAGRDSLEAEQVMPFMNTVIMRSVLHGRKDVMSQIQATISDNHOF 10098  
QY 459 ----- 458  
| | | |  
Db 10099 FLRQIOAACQGVQSTQNDADFDSLFTFORRPDTTSTKTMLYESVNGASEVEYPPVAVE 10158  
QY 459 ----- 458  
| | | |  
Db 10159 AEIVDGSILWRIAGKSSAFDEEGVNRNILDLTLEDVLQAIVTQPHPTLAFDNQEVSI CDLP 10218  
QY 459 ----- 458  
| | | |  
Db 10219 AFQOQEKTSQTAGVSDSEIVEDEEHWSDTESVRAEIAKVTKTPVQEISKTASIQNLGV 10278  
QY 459 -----SLW----- 461  
| | | |  
Db 10279 DSINAIKISALLRQKIIIRVTVSEIVRAGSVWKIASVAQNKKQKQKQAKFADAKVIVDL 10338  
QY 462 ----- 461  
| | | |  
Db 10339 MAQKFTPDKGYEDQHQIEQILPATAGQVYMLGAWQTNGQLFFGQFEYVTAVTMTDDI 10398  
QY 462 ----- 461  
| | | |  
Db 10399 KOAQKVTASNAVLRTVVFATQDNTEPVLQLVLRLDAPTSFVDLDNEETPASIIIGQPFVRL 10458  
QY 462 ----- 461  
| | | |  
Db 10459 TAQKVDDTYKLCIKIHHALYDAISLSLQELGNLYLQDSATTQOSPVKPADFTIAQPLGTV 10518  
QY 462 -----WWL 464  
| | | |  
Db 10519 DKTAKPFW 10527

RESULT 5

T31307  
type I fatty acid synthase homolog - Cryptosporidium parvum  
C;Species: Cryptosporidium parvum  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T31307  
R;Zhu, G.; Marchewka, M.J.; Woods, K.M.; Upton, S.J.; Keithly, J.S.  
submitted to the EMBL Data Library, August 1998  
A;Description: Characterization of a type I FAS gene in the parasitic protozoan Cryptosp.  
A;Reference number: Z20993  
A;Accession: T31307  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-8243 <ZHU>  
A;Cross-references: UNIPROT:O96554; EMBL:AF082993; NID:g4092068; PID:g4092069; PIDN:AAc9;  
C;Genetics:  
C;Note: FAS1  
C;Keywords: carrier protein  
F;100-714/Domain: acetate-CoA ligase homology <ACL>  
F;791-861/Domain: acyl carrier protein homology <ACP1>  
F;906-1308/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>



F:2924-2992/Domain: acyl carrier protein homology <ACP2>  
F:3062-3463/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>  
F:5157-5227/Domain: acyl carrier protein homology <ACP3>  
F:5274-5689/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>

Query Match 21.2%; Score 520; DB 2; Length 8243;  
Best Local Similarity 2.7%; Pred. No. 0.17; 137; Indels 7589; Gaps 70;  
Matches 217; Conservative 100; Mismatches

```
QY 3 RKEMVRDVPKM-----FVLISISFL----- 23
DB 197 RKEEYINEVQKAINDIILNDSNHDFLGFDFNINIVAFLLQFTSGTSQPKGVMTGHNLL 256
QY 24 ----- 23
DB 257 YNIIHICISSYSPSVLEBHYKNELDISIPAGITVLKKNELTEDELSFIHNEFCEKNIMSL 316
QY 24 -----LVSRI----- 28
DB 317 NDIKYDKLYLMSIINKVLYKNTWSSASVFWLFPVYHDMGLIGFCTPLPFGCNIQMS 376
QY 29 -----NCKV----- 32
DB 377 IDFIKPYLWQCMDKYNCSVGAPNFAFEVVVRKTPKEILNQLNKLKHVFALLSGAEP 436
QY 33 ----- 32
DB 437 KTTIDRFEAFKSVGIKHNIKPAYGLAEHTLIVAGSNSFOQEVNHTVNTKLRKNIV 496
QY 33 ----- 32
DB 497 BIKEAKESSTETTSFVSGGIYVYKIDIRIVNPESLKEVTPGNVGEIWISSSVTLGY 556
QY 33 -----MSKALYNR----- 40
DB 557 NKEETKTFNANFTMLDCKTSKTYWRTGDSGFINDMLYISGRKIDMIIRGNFYQN 616
QY 41 ----- 40
DB 617 IEEVIDGVGRQSVAVSVTQADGEAIGVAVEIRMETSIILGRVRRFFKEPAYDNIV 676
QY 41 -----PW----- 42
DB 677 AISKAVFTGHLGVHYIWLSPRTILKTSKIRRSQTRDAIPSKLPLFEWCEQSPM 736
QY 43 -----RGLVLSKIG-----KYKLDOLKLE 61
DB 737 SDILVERSSQFISPTSMIISGVQNKSFYEKTHSGQOKISBQANDEKNEFDIDQVK 796
QY 62 ILR----- 64
DB 797 IIEFAAQVLGITQKDLNAPLHEYGIDSLGSIIRLSEMMNEBFGVELDSTLLFNVP 856
QY 65 ----- 64
DB 857 IVDVFAQISGKELHRKLNLRSGIDSYSDIWVVGMSCFPGSGSTPGEFWSLLQSGV 916
QY 65 -----QLETTISTK----- 73
DB 917 IGEIPKSRWNIDEYVSTDLLENKMYTKEGGFENIDQFGASPFKISHAEARSMDPQ 976
QY 74 ----- 73
DB 977 FLEKSFEALKDVGTYITDLTKRNISVPAGCCSNWDTHVCKSEANSNIGTYAATSHA 1036
QY 74 ----- 73
DB 1037 ANRVSYTLGLAGSSVTVDACSASIVLHVSLOEISTQCBAAVVGGINLMSPOITVAL 1096
QY 74 ----- 73
DB 1097 CKARLMSMDCRCKSFDAANGYVRGEGVGLVLTMTMENANNGEKIYSVIKSGSSV 1156
```

```
QY 74 ----- 73
DB 1157 ASLTAPNGISQONVINSAGLEAGIRPNDIGYIEAHGTCTSLGDPFIEVSAKLSIFSR 1216
QY 74 ----- 73
DB 1217 GKPLIIGAVKTNIGHLEGAAGMAGIFKVLISLINDTVPPNLFHKNINPHINLKGPP 1276
QY 74 ----- 73
DB 1277 TSNIALQTYDGSKTCAGVSARFGGTNAHAIFEKPNLSIDLPIKTVKIETNKARNPV 1336
QY 74 ----- 73
DB 1337 VFGQGFQHVNMKGYFEREIVYREWFLLCKCKIIDPFLKISLKLVLVYPETKMEK 1396
QY 74 ----- 73
DB 1397 KILQBEYAQCALPAVQVSLAKLLESKGIYPGAVIGLSDGELVASVYCGSINLEDGL 1456
QY 74 -----YNVSK----- 78
DB 1457 MLRASLLSEFSAGSGKMAFPVSKEDAEMALADVGPQHNNRVISIFASISSNQVLLS 1516
QY 79 -----QPVKQLTMTNTEF----- 90
DB 1517 KLETVPFILREYYPANAHSIDAKPLRTKYSGFTGNGMGLIEPISNLMENIALET 1576
QY 91 -----POYV----- 94
DB 1577 ISGTSGYFVVEELTSPKFWGQNLNKLISKALTALINGASTILDLMPIRSSEFEISE 1636
QY 95 ----- 94
DB 1637 FNPENIDKDTIYSLPBDTNLESYEPNKFNMWIKVSVSILLKNPVNVEDACDQIINS 1696
QY 95 ----- 94
DB 1697 TYSRESFPWNSFKHSILGEQKSIDRSVKIVSTLNNKSIKLFSDHKVYETIIVIPG 1756
QY 95 -----ILAG----- 98
DB 1757 AAASIIIVSMDDSILOGSNALFGESIAVKLNVLFEKPVVNVNRTSLSESSECTEVI 1816
QY 99 ----- 98
DB 1817 SYNLANQKTNLICNIESDGSISIOYCNIDNQDFDYEADNSMIDCFSCAKISEGEF 1876
QY 99 ----- 98
DB 1877 LSEDLAQKNNTIKEDPKIMYEYEBIGLYGKKFRVVKIYRNETGSIAGLKIYLPET 1936
QY 99 -----PI----- 100
DB 1937 CKFOTPSGFFHPAILDGLAHVAGSLLYKDKSPKQAMVPVSEIEIEIKYVNNQMIW 1996
QY 101 -----QWYSI----- 105
DB 1997 AAAYLSEATQNSIFDLVDFDSEGCIGLEKVSRLRQFNKNISRPVTVSHNEILLW 2056
QY 106 ----- 105
DB 2057 KSDVLKPSKNISKESDPLKIALIHSGICNISEKODIQLPKANISFVNLDKIKESN 2116
QY 106 ----- 105
DB 2117 ERLDITSLDSLVELSPVELALSAKIDCEKISIDILEYVMKICKLYKLKQAKPTQ 2176
QY 106 ----- 105
DB 2177 KINIPFWIITTSQNLIYNINENICIPNNSGLWGLAKSANLEISSLLTNFSQPIK 2236
QY 106 ----- 105
```

Db	2237	LEITDFSDNQYLKRTIGSLVRWIGNFEYSSSLKEKENLIFKFEREFSITSKNIYVNRINKP	2239
Qy	106	-----	105
Db	2297	HKQIVGPVELFMQNRGSIISNLIESEPIQLEKNDYEVNVRVKSIGLFRDVLNVMGly	2356
Qy	106	-----TV-----	107
Db	2357	PGDPGPGDCSGTVVAVGEGVHKIKVGDNVFGIAPGCLKTYVTDSNLLCKIPKGTPE	2416
Qy	108	-----	107
Db	2417	QAAALPVVATTVESYSLKDIAIINKGDIVLHVAVTGGVGLMVVQYKAIKAVYGTAGES	2476
Qy	108	-----	107
Db	2477	KVEYALSGVERVSSRNADKFEDMKDLEKVDIVINSLIEDFIPNSLNLKKGCGFVE	2536
Qy	108	-----LW-----	109
Db	2537	LKRGIWTEEBEKEKRPDIKYKCVADVMMEEEDPSWFGMLTRIKNLAEGBIIESPLKV	2596
Qy	110	-----	109
Db	2597	PMRGSDENGIDAFRYMQRAHQHIGKVIKIPYLAGIGDINSIPBECLITGSGSLGLVITE	2656
Qy	110	-----	109
Db	2657	TLNEGITNISIIISRGIESQSDSIKAKWEKFKRILNNKNGSGIRFFNIDIGNKDSVY	2716
Qy	110	-----	109
Db	2717	KLINTEFKNLKHIPLHAGLVSDSNIETQTRESIEEVFKPAIGAWNLHNCBELGLNKEL	2776
Qy	110	-----	109
Db	2777	KTFVYSIASGLGNFGQTNYSAA NSCLDSLSQYRVSKGLSSRSIQWGAWDEGMAWNR	2836
Qy	110	-----FDFVSTQL-----RK-----PAKY-----VYS-----	126
Db	2837	QHLGLVMSPISKTTGQOVFDLITTOIYEENRKSATIMCLPKWKTYINTIYSGENIPQ	2896
Qy	127	-----	126
Db	2897	FSLVNTNLRKQONTILLKMSKEERTVIEKQIISLSKSIILGVEIKSLDQPLQDIGDS	2956
Qy	127	-----	126
Db	2957	LAALFERNALSKKFNIKLSATTLFDYPTIRGIKDHISEKLSVKESIYETNKTQISNLIS	3016
Qy	127	-----	126
Db	3017	QDNLSDDNVGIAGLACRPFPGCGNTPDTLWKSILIRGYSINPLSGRMFNSVGAFASREI	3076
Qy	127	-----QYNH-----	130
Db	3077	PIQRWNHNLVYNADPKTKGKYSYKACFIDSIDMPDNKFGITNIEAKHMPQQRIVLET	3136
Qy	131	-----	130
Db	3137	CYEALVAAKIBETLIGFQMGVPIGCCSNDWSFLQSRKGMAPPFTGTGAANTTISNRVSIV	3196
Qy	131	-----TAKTI-----TFRPPPC-----	142
Db	3197	FGMRGFSMTVDITACASSLTAACIAMDSFRNKECNGAIVGGVALLSPNLPIAFCKARMLS	3256
Qy	143	-----	145
Db	3257	VDGKCKTPDASADGYVRGEGCGAIVILKQSEOKKONTPIILGRIKGWGNHVRGSASLTAP	3316
Qy	146	-----PSMTCLSEM-LNVSKRNDT-----	163

Db	3317	NGPAQTCVIMALNOSKLNSSDIDYETHGTGTALGDPIELGALKSVFGRNADYKRS	TPL	3376	
Qy	164	-----	-----	163	
Db	3377	VLGALKSNIGHLEAGAAGIAGLKLVLKHETAIRIAHLTKINPHLDEDFNVIIPQ	KTV	3436	
Qy	164	-----	-----	168	
Db	3437	SLKCNKNLIGVSGFGGCGNTHIVIVESTEKNSEDIAADFKNHILPVFTGQGSQ	YI	3496	
Qy	169	-----	-----	170	
Db	3497	NMCKGIYETEPFRENMLCNEIVSKIINISLLDIYPIDGNKDAEFKANMLNDR	YAO	3556	
Qy	171	-----	-----	170	
Db	3557	PAIPAVEYSLAQLWISKIKPDPSVLGHSGLGEFIAATISQVMTTIEDALT	VTHRASIMAST	3616	
Qy	171	-----	-----	170	
Db	3617	PALDGMVACRLTESQVDDTIKRFGLBEESASLAUVNGOKSVTISGKESVYK	EYKILEYNNVG	3676	
Qy	171	-----	-----	170	
Db	3677	SRFRDLVSHAFSPLVSEASDKPKLLENILKKTVEFISVTGKKEIS	STAKYWS	3736	
Qy	171	-----	-----	170	
Db	3737	KHILNTVRLSDAVHATIESNSRMTFIBITSPKLSQLLKALIPVNSSHIS	VKCTCKLVN	3796	
Qy	171	-----	-----	180	
Db	3797	SEIDFIOLEBQFEGSENSFTNHANNLDDSWRFKLKRRQISWDISHPIAP	LHDP	3856	
Qy	181	-----	-----	180	
Db	3857	NDIEWKENISISKNARELVEKLEFTCMISPELLDFDNHKVLQGSILPGA	AFVDFMAT	3916	
Qy	181	-----	-----	184	
Db	3917	VALNVTKSQFTGLVIGAMPDWIQKGFPRNPFI	MSAKQYKYLNTQFDSQYSD	EDNSS	3976
Qy	185	-----	-----	184	
Db	3977	RSSDFNIIMSGKNNCOISIESLKMDEETVWYATCDEVYVSSIEAQT	SLCIISEWPSS	4036	
Qy	185	-----	-----	184	
Db	4037	ELISNKKVQEDIEYKMSIAGLOYGSRFTLKENWKVTSNSAVGIIQ	SELFTKNCASP	4096	
Qy	185	-----	-----	184	
Db	4097	DNLAVYESIMNERGFTIHTLLDGVLHMCASILGQSDKETFNELHKEAT	WVPISINKCL	4156	
Qy	185	-----	-----	184	
Db	4157	ITSKIDTSKYI	GEKWTAFVQLKSSDKOSAIWNVALKSMGVPIAVLLGVSLRSVK	NGVI	4216
Qy	185	-----	-----	199	
Db	4217	STHTIKHVIPNELLWRIDWDEVISITNSEIVSLPTKVCN	TENESVQGINANLVEKRK	4276	
Qy	200	-----	-----	203	
Db	4277	VLLTNSCNISKELLYSEYFTADIVCEASKLSLAEIAT	IYGRDQELPPSPSNEAKYDL	4336	
Qy	204	-----	-----	203	
Db	4337	ICLHIPSNDKNAIVSLLPVLVKILRSMFENKERLIPVR	IITENIPNVDGMNLEFRT	4396	
Qy	204	-----	-----	214	
Db	4397	NSGVAPFIKARQELFSDQSVSLGLDIFETABDLNKA	ILISFVKDILCFKKREDITSN	4456	

QY 215 -----AQR-----NCTHSFYLWN----- 227  
Db 4457 GTLTESQRLFLNTEYEPELVLKKGKEDNCEAHYRIITPKVLGRGEAALIGSKLNL 4516  
QY 228 ----- 227  
Db 4517 SSRGAINNLELLPLSNEERKCPESANTVIRVRSIGLNFDRVLNMGLYPGDPGPGDCS 4576  
QY 228 ----- 227  
Db 4577 GTTVAVGEGVKHIVKVDNVFGIAPGCLKTYVTIDSNLLCKIPKGTPEQAAALPWVATTV 4636  
QY 228 -----AMSRNLFR 235  
Db 4637 EYSLRDIANIKKGDIVLVHAVTGGVGLMVQVCKAIGAKVGTAGSKVEYALNGVER 4696  
QY 236 V----- 236  
Db 4697 VSSRNADKFKEDMKDLEGKVDIVINSLEIDFIPNSLNLKKGCFVELGKRGWTEBEM 4756  
QY 237 -----PKYING-TKLKN----- 247  
Db 4757 KEKRPDIKYCVANDVMMEDPWFEGMLTRIKNLVEEGIIIESLPLKYVDMRGSDENGID 4816  
QY 248 ----- 247  
Db 4817 AFRYMQRAHQHIGKVIKIPTPPKYLELDTENHKLNNVSEKEDRHNNVEYGIHVITGGIGG 4876  
QY 248 ----- 247  
Db 4877 IGRIIWKMLBEGVRKIAILSRSANNDLSRDIPEIKOYLNTDIIQIECIKCDVSILSQVE 4936  
QY 248 -----TMRKLKKEQAP----- 258  
Db 4937 SAFREILVKFGETNQIHGIHRAAGILMDGATASQTMEMESVYAPKYVGAWNLHECCBKF 4996  
QY 259 ----- 258  
Db 4997 DLNKLKHFVMPSSVASLGNFGQNTYSAANCLDSLVEYRRNKGMCOTSIQGWPIEQG 5056  
QY 259 -----VKEQPE----- 264  
Db 5057 MAMNIKQHLKLVGMHGISNEMGIRVINDVILYQNTDLDKNSLGYPVITCQSFNNVYMR 5116  
QY 265 ----- 264  
Db 5117 QLYYNSKYPSPFKDVVASMSTGEVKNIGMDFRGMSEVYSYVKDSVHNIIVEQIIGIQSS 5176  
QY 265 ----- 264  
Db 5177 TLDLTPSELGMDLSAIELRNALISRFGIKISSITALFDYPTLRSIIQHLTDSLIVSET 5236  
QY 265 ----- 264  
Db 5237 RDAFSDLGMSISEREYSVAIIIGVSCRLPGRVKSFNLDYEIVLLKSKOCITEPLDRWPPD 5296  
QY 265 ----- 264  
Db 5297 LYDPDPNRSKSVYNQGAFTDIDLFDNTPFGLSSTEATNVDPQKYMLEVAYEAMVNS 5356  
QY 265 -----KKAKKT----- 270  
Db 5357 GITKRLSITSIKTAKTGDDPVLGMVFIGCCNTDWHFLQAKGFENFTSYTSGSGASSLVS 5416  
QY 271 ----- 270  
Db 5417 NRVSYSFGRGFSQTI DTACSSSLVAMDAALKLKGSCHTALVGGVNLMLSPHLFVAF 5476  
QY 271 ----- 270  
Db 5477 RARMLSPDCRCKTFDASANGYVRGEGGAILLTRSEIGGKIRPIAYVLGSATNHNRS 5536

QY 271 ----- 270  
Db 5537 ASLTAPNGPSQTEVIQNALRNARLTPNDIDYLESHGTGTPLGDPIEFGALKTVFGKKXI 5596  
QY 271 ----- 270  
Db 5597 KRSQPLILGALKTNIGHLEGAAGVGLLKLILVLKNRIAPKILHFKOKLNPHIDTNEFDVE 5656  
QY 271 -----QSTTTPY--FSYT-- 281  
Db 5657 PPTELKPTIISNKSXLVGVSSFGGCSNHSVILESSANELNTESSQKNPYLLIFAYTG 5716  
QY 282 -----TSA 285  
Db 5717 QGCOYINMKQLYDTEEVFRNEMDKCNELLYPILGRSLISTIYCEKTMESKLNHTSIS 5776  
QY 286 LNVTTNTYS-----ITTAARV----- 303  
Db 5777 QPAIFSIYSLTKLWESKGI VPSAVMGHSLGEFAAAVTSGLCLKDAITLVARRAEIMAS 5836  
QY 304 -----STSTIA----- 309  
Db 5837 LPQNDGIMVACRVSEDQVLDISINKLNLKDSAAVAALNGPKSVTISGSRSSVMQVLNLMG 5896  
QY 310 -----YRPDSSPMKSI----- 320  
Db 5897 EDRYKQLDVSHAFPLSVSSABEKPEKVIENIQFPEPKLTPFSSVTGNIESKLLIENKYW 5956  
QY 321 -----MATQRLDA----- 329  
Db 5957 AQHIIHVVRYDALETAISTDEENIAIEIGPKSVLTSMGKQAVGGIOLVOSSNENISNN 6016  
QY 330 -----TWV----- 332  
Db 6017 ELTWIPTIDNSKNTIDFNDALKAAREYIKYISSEVCVHKWHSRPHWADLTSVHPLGK 6076  
QY 333 -----YTT----- 335  
Db 6077 FYSSISDENLAYSTDMLSKSVFKNIVEDHVIPTDNIIPAAALIEASACAAPKQSRSEKF 6136  
QY 336 ----- 335  
Db 6137 PSKYVENDSMIDIEKPLVSNQDETTSEMEVFVAKGAIQIRKCISEAFYEFEFSGPQII 6196  
QY 336 ----- 335  
Db 6197 ANDTVFSNRIYHHASISNSRILSKDEVDKQCFPNLVDYLEKYSKDTMEVPIDNIYQM 6256  
QY 336 -----LRYQN----- 341  
Db 6257 NEIGLQYQKNYQCLKTVKVKIDNSDEKKIKSGSNQINEVFCYVXRPISDFDLTELRIH 6316  
QY 342 ----- 341  
Db 6317 PGMLDSVFQASSVLFADELLEGKDTTKNKRGHVAMAPIGFTKCFYGRIOHNSDVMGLVK 6376  
QY 342 ----- 341  
Db 6377 LNRPEKKGNTAFIDVTILWDSKTNILAAHFQDILLKGFNSAASLLSQEVRIPNQPLWRTYN 6436  
QY 342 ----- 341  
Db 6437 KKI VDRNTYIENSSTPFEQENKEDQDKBFNSEDSDHDNQIQIKILIFGIPTVIKELICSQWD 6496  
QY 342 ----- 341  
Db 6497 NKYKISTTFFICLEDLTIDLIDOKIQEENWSAMFYLGLMDNLSSKQVMGDILKISQVI 6556  
QY 342 -----PF-----CE----- 345  
Db 6557 STISVSITTAFFVCLTSFNDCEEDGNIPRHSGIKGFVKRTARVELENVIGKSHINISHIECN 6616  
QY 346 ----- 345

Db 6617 SSLVNDPMTLRTISYIASIEGKIEBITLERDLTISESGIYAPRLSINFPVFGAIKL 6676  
QY 346 ----- 345  
Db 6677 TMSRGAIGNLKPKPQPSERVPPANTVEIRVASIGLFRDVLNVMGLYPGDPGGD 6736  
QY 346 ----- 345  
Db 6737 CSGTVVAGGVKHKIUGDNVFGIAPGCLKTYVITDSNLLCKPKGTFPEQAALPVVAT 6796  
QY 346 ----- 345  
Db 6797 TGEYSLRDIANIKKGDIVLHAVTGGVGLMVVQYCKAIGAKVYGTAGSEKVEYALNGV 6856  
QY 346 ----PSENRTAVSEFMKXTH----- 361  
Db 6857 ERVSSSRNADKFKEDMKDLGKVDIVINSIEDFIPNSMLLKKGGCFVELGKRGIWTEE 6916  
QY 362 ----- 361  
Db 6917 EMKEKRPDIKVCVAVDVMEEDPSWFGMLTRIKNLVEEGIIESLPLKYDMRGSDENG 6976  
QY 362 ----VLIRNETPY----- 370  
Db 6977 IDAFRYMORAQHGKVIKIPTPKYLELDTENHKLNVHGSEKEDRHNNYEYGIHVITGGL 7036  
QY 371 ----- 370  
Db 7037 GGIGRIIWKMLEEGRKTAIILSRANSNDLSRDIPEIKDYLNVDIIQIECIKCDVSIILSQ 7096  
QY 371 -----TIV----- 373  
Db 7097 VESAFREILVKFGETNQIHGIFHAAGILMDGALASQTMMESVYAPKVYGANLHECCE 7156  
QY 374 ----- 373  
Db 7157 KPELNKLNKHFVWSSVASLLGNFGQTNYSAA NSCLDSIVRYRRNKGCGTSIQMGPWIE 7216  
QY 374 ----- 373  
Db 7217 QGMAANLKOHLKVMGYGINSNELGIRVLNDILCFVSKSSVIGCOSLKWDTFMRRYNTIP 7276  
QY 374 ----GTLDMSSL----- 381  
Db 7277 TFLSEVDYSSQGRSQGRVDISKLSDEELVWVIAQVAAQCASPTLPSPDTVLLDGLDSL 7336  
QY 382 -----YYNETMFVE-----NKTASDSNK----- 399  
Db 7337 GAVEFRNSVLEMTGVKLPQTIVFENPTIYATSMYVRDQNSGTSKDSKSHSQTSNFQAK 7396  
QY 400 ----- 399  
Db 7397 MTIEQWLLSSLKQSERIYIAEAFKKYQTI SNLTETDIISALEELGVENNOYDILLYV 7456  
QY 400 -----TTPSPS----- 406  
Db 7457 SNNELISSEBNSKONDSNKKLEQELIFDPIADVEIKNSMKLDLDGISTTDPSEFK 7516  
QY 407 ----- 406  
Db 7517 TALLTGVTVFVGRILLVLLIEBFKNMQIVCLVRASSEKGLERINVCERAEVWNPTYAS 7576  
QY 407 ----- 406  
Db 7577 RIIVECGNFEERYLGLSEERYELCKEIDVVYHIGDVLNLLSNYKRLKNTILSLVGIIN 7636  
QY 407 ----- 406  
Db 7637 FCTIKLKLHFSSTLGQFPAPFAVFTREPENAVVKETEGPSTREMSRLPPTTRQGYPWS 7696  
QY 407 ----- 406

Db 7697 KWAEBQILETAHQOGLPLSLIYRLPNTYIASDTGYTNKTDYATALLIASILEGMRPFGSST 7756  
QY 407 ----- 406  
Db 7757 APLTPVNTICEIIISASKKKERKWRYNLLDTRILGKKGHEAWASQLGLGNYKGVAIDEF 7816  
QY 407 -----MGF 409  
Db 7817 FSAIKSRGPSPFIPKFPVPLMQYWRHYWFDNTERQPPFPDTSNVSEDMPEISWPPLEITP 7876  
QY 410 QRTFT----- 414  
Db 7877 QNSFLYCAKRGYFPSPNSSSISFTPIKCYSDALNEIKAELSLRNVNFNQNKYDFICRKMOK 7936  
QY 415 --DPLWDYL-----DSLL----- 425  
Db 7937 NADSIYDSLKKVSLNFCNFKYSYIILLKQSVQVYIQUESLNDLSILEKFMQMDCYFITGTS 7996  
QY 426 -----F 426  
Db 7997 FEILELRVIRKLLNNYNELKFVDIACPYXPTKNSRRLILEQLKIATSIPLKRFPAVD 8056  
QY 427 LDEI-----RNF----- 433  
Db 8057 LDEILTDSDILLEMFPAPFTPLIYKGYFEEYKKNKITGCSEVMEAYSSFQNYIIQTLI 8116  
QY 434 -----SLRSPTY 440  
Db 8117 KKNYNTBYPIIFSSPFFLFPFVKELQIFENSKIIVVDNKPNDSTIQGKLSMWKSLKS-KY 8175  
QY 441 VNLTPPE-----HRAVN-----L 454  
Db 8176 TNLDSVEENSEVYLLDILEMKHRFVINEPSIDKNRLIYLKDFGEIHNKISVOLQDFC 8235  
QY 455 STS 457  
Db 8236 STS 8238  
RESULT 6  
T13931  
projectin - fruit fly (Drosophila melanogaster) (fragment)  
C;Species: Drosophila melanogaster  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T13931  
E;Daley, J.; Southgate, R.; Ayme-Southgate, A.  
J. Mol. Biol. 279, 201-210, 1998  
A;Title: Structure of the Drosophila projectin protein: isoforms and implication for pro  
A;Reference number: Z17815; MUID:98300339; PMID:9636710  
A;Accession: T13931  
A;Status: preliminary; translated from GB/EMBL/DBBJ  
A;Molecule type: DNA  
A;Residues: 1-6658 <DAL>  
A;Cross-references: UNIPROT:O76281; EMBL:AF047475; NID:G3337430; PID:G3337431; PIDN:AAC2  
C;Genetics:  
A;Gene: projectin  
A;Cross-references: FlyBase:FBgn0005666  
A;Map position: 4  
A;Note: intron positions not resolved (incomplete sequence)  
C;Keywords: muscle  
Query Match 20.4%; Score 500; DB 2; Length 6658;  
Best Local Similarity 3.1%; Pred. No. 0.28; Mismatches 130; Indels 6187; Gaps 68;  
Matches 208; Conservative 98;  
QY 5 EMVVR-----DVPKMFVLISISFLLVSPINCKVMKALYRNPW---RGL----- 45  
Db 1 EFRVAVNRAGPGEPSDKTK-----SIIAKCRFVKPFIVGEGLNKVTV 43  
QY 46 -----VLSKIKY 53  
Db 44 KKGQIRFDIKVDGEPPEAATWKGTDNLKFDNORICLDQLERNSSITIKKSVRDTGKY 103

QY 54 KL----- 55  
Db 104 KLVLNSGGTIESBAQVVLDRPLPPGGFPPEBIRASHIMKWKRPDDGGCBISGYAL 163  
QY 56 ----- 55  
Db 164 BRMBETGRWTPAGEVGNETSFDPKGLTPNKKYKPRVKAINKEGESEPLETFDASCCKNP 223  
QY 56 ----- 55  
Db 224 YDPPSPSQPVIDDYDNKSVLLKWKRPDSDGCRPITHYIVBEIKDKFAPSWSEVAKTDDPN 283  
QY 56 ----- 55  
Db 284 PECNVEGLKERWYQFRVRAVNKAGSPSPQTDNLHCKHKNLKPQIDRSTFKRVTIKSG 343  
QY 56 -----DOLKLE-----ILR----- 64  
Db 344 RTHKWSVDVLGEPIPELHWSRRDDIPLTNGDRIKIENVDYHTDFSITNVLKDSGFYTLK 403  
QY 65 ----- 64  
Db 404 AENRNGIDRETVELVLGKPSKPGPLAVSDVTASGSKLQWKPEDEVGGVPIKEYVVEKM 463  
QY 65 ----- 64  
Db 464 DTATGKWRVGRSPGEKEPPSFDVTGLSLGSEYMFVRSVAVNEEGDAEPLTTLVGVAKDP 523  
QY 65 -----Q 65  
Db 524 PDEBNKGTPEVTDYDNQISLKWAAPNNDGGAPIQKYIIEKKNKNTWEKALEIFGDQ 583  
QY 66 LETTIS-----TKY----- 74  
Db 584 LEATVAGLOEYGEYQFRVIAVNKAGLSPSPDASVPQIVKYKKLAPRIDRSNLKPELLIRAG 643  
QY 75 -----NVSKQP-----VNKLTWNTFEP-----QYY 94  
Db 644 KPIRYDVNVRGEPAPVITWYQNDKELKPELPSSSEIKNIPYNTKISIIETVRKHTGIYK 703  
QY 95 ILA----- 97  
Db 704 IIAVNEHQDEATVEVNI LAPSPKRGPLDVKQVTKDSCFKWKPKBEDDGGKPI SAYQVE 763  
QY 98 ----- 97  
Db 764 KFDKKGRWVPLGRTSANDTEFDVKGLOEGHEYQFRVKAINEEGESDPLSDSDSIIAKNP 823  
QY 98 -----GPIQNYSI-----TYLWPDFYSTQ----- 116  
Db 824 YDAASKPGTPTNVDYNEHVMVKLWEAPRSDGGAPISGVYIEKKDKFSPINDEILSTNTSV 883  
QY 117 ----- 116  
Db 884 PEATVEGLVGNIIYQFRVRAVNKAGSPDSDATEPHLAKPRNLKPYINRDKMKPIKVRAG 943  
QY 117 ----- 116  
Db 944 QPVKFDVDVKGEPAAPSLTWFLKETELTSTGQVRLENIDYNTKLTLLDTRKQSQYKLR 1003  
QY 117 -----LRKPAK-----YVYSQY- 128  
Db 1004 ENINGVDBAVVEVIILODKSPKEGPLEVSDIHKEGCKLWKRPKDDGIGIPITGVVIEKMD 1063  
QY 129 -----NH-----TAKT----- 134  
Db 1064 TATGKWPAGSDPEKYDIEIKGLDPNHRVQFRVKAINEEGESEPLETESAITAKNPFV 1123  
QY 135 -----ITRPP----- 140  
Db 1124 SAPPLPELEDWDEHHVKLWEPPIRDGGSPITWYIIEVMDKDPGEFVKA VETDSPVCKG 1183  
QY 141 ----- 140

Db 1184 VVKLEBGOQYKFRVRAVNKAGSPDSEQTNHVAKPRFLKPHIDRVNLKPVIVKTGLSI 1243  
QY 141 ----- 140  
Db 1244 SLDINIRGEPAPKVEWFFNNSSVTSDEHSVKIDNDVYNTKFFVMRAORSQSQKYIITKATN 1303  
QY 141 ----- 140  
Db 1304 EVGDEAELEVTVLGKPKPKPLQVNDITKHSCKLKWPKDDGGSPDIYIEKLDPH 1363  
QY 141 -----PCGR----- 144  
Db 1364 TGQMLPCGKSTPEAKVIGLHKGKAYFRVRAVNKEGESEDELETEKPIIAKNPVDPEPRP 1423  
QY 145 -----VPSMTC----- 150  
Db 1424 GKPEPTWMDKDFVDLAWDPKNDGGAPIQKYVIQMRDKSRAWVDSATVPDCKNGVTG 1483  
QY 151 -----LSEMLN----- 156  
Db 1484 VEEGHEYEPRIVAVNAGSPSDSVKSIAKPRFLKPHIDRKNLQKKIMRSGQMLHIDA 1543  
QY 157 -----VSKNDTG----- 164  
Db 1544 LIKAEPPAKVTWYNTKTEITKTSDHIKIENEDYKTFIMPVKRADRGYIIVTAKNDSGSD 1603  
QY 165 ----- 164  
Db 1604 TVEVEVLCKPKPKPLAVSNVTAETLHLKWKPEDDGGDPIEQVLVERMDTETGRWV 1663  
QY 165 ----- 164  
Db 1664 PVLTTKTPEADVTLGTEGKEYLFRVKAINEEGESEPLVTDIPTKAKNPFDAADTPGKQI 1723  
QY 165 ----- 164  
Db 1724 VDMGNNHCDLWRAPEDDGGASITGYIVERKDPNTGKWKQALETSTPDCKARVNDLIAGN 1783  
QY 165 ----- 164  
Db 1784 KYQFRIMAVNAGSKSPSBDQMTAKDRAPPKIDRTNIKDITSKAGQHIRFDIKVSGE 1843  
QY 165 ----- 164  
Db 1844 PPATKVLHKNARLENDSDSNYNDMESYRTKLTVPISKRPHSGKYTLKAENESGRDEASF 1903  
QY 165 -----EQCG----- 169  
Db 1904 EVIIVLDKPGPEGLRVTDVHKEGCKLWNAPLDDGGLPIDHYIIEKMDVESGRWLPSGR 1963  
QY 170 ----- 169  
Db 1964 FKESFABLNNLSPSHYKFRVLANTEGESEPLTGESQVIAKNPFDEPGKGTPEAVDWD 2023  
QY 170 ----- 169  
Db 2024 KDHVDLWVRPPINDGGSPITGYVVEKREKGTDKWIKGTEITIPCLGEECKATVPTLNENC 2083  
QY 170 -----NFTFNPMFF----- 179  
Db 2084 EYBFRVKAINAAGPEPSDASKPIITKPRKLAPTILDTKNIIRTYNFKSGEPIFLDINIS 2143  
QY 180 -----NVPRWNTKLY----- 190  
Db 2144 GEPAPDVTWNNKSVOTTSFSHIENLP-YNTK-YINNPERKDTGLYKISAHNPFYQDQ 2201  
QY 191 -----GPTKV----- 195  
Db 2202 VEFQINIITKPGEGGLEVEVHKGCKLWKWKPKDDGGEPVESYLVERKDPDTGIWLP 2261  
QY 196 -----NVD-----SQTIFYLG----- 206  
|||

Db 2262 VGRSDGPEYNVDGLVPGHDKFRVKAHVKNKGESEPLETIGSIIAKDPFSVPTKPGVPEPT 2321  
QY 207 ----- 206  
Db 2322 DWTANKVELAMPEPASDGGSPIOGYIVEVKOKYSPLEWEKALETNSPTTATVQGLIEGNE 2381  
QY 207 ----- 206  
Db 2382 YQFRVVALNKNLSEPSDSKIFTAKKRYIAPKIDRNLRNITLSSGTALKLDANITQOP 2441  
QY 207 -----LTALLRYAQRNCTHSFYLVNA----- 228  
Db 2442 APKVEWKLNVHLOSGKNVTIETPDYITKLVIRPPQRTDSGE-YLVATNTSGKDSVLVN 2500  
QY 229 ----- 228  
Db 2501 VVITDKPPNGPLQISDVHKEGCHLKWKRPSHDGCTPIEFQIDKLEPETGCWIPSCRS 2560  
QY 229 -----MSRNLFRVP----- 237  
Db 2561 TAPQVDVTGLSPGNEYKFRVSAVNAEGESQPLVGDESIVARNPDPDPKPNLKAATDWDK 2620  
QY 238 -----KY----- 239  
Db 2621 DHVDLAWTPPIDGGSPISCIYIEKQDKYKWERALDVPADQCKATIPDLVEGQTYKFRV 2680  
QY 240 -----INGTKLKNMTM 249  
Db 2681 SAVNAAGTEPSDSTPPIIAKARNKPPIIDRSSLVEVRIKAGQSFTFDCKVSGSEPAPOTK 2740  
QY 250 RKLKFKQ----- 256  
Db 2741 WLLKKKEVYSKDNVKNVTNDYNTKLKNSATRSDSGIYTVFAENANGEDSADVKVTVIDK 2800  
QY 257 -----APV-- 259  
Db 2801 PAPPNGPLKVDIENSESCTLHWNPPDDGGOPIDNYVVGKLDETTGRWMTAGETDGPVTA 2860  
QY 260 -----KEOPEKAKKTQSTTTPYFSYTSAAI----- 286  
Db 2861 LKVGGLTPGHYKFRVRAKNRQGTSEP--LTTAQAIIAKNPFDVPTKGTPTTKDPEKE 2917  
QY 287 -----NVTNTV----- 293  
Db 2918 FVDLEWTRPEADGGSPITGYVVEKDKFSPDWEKCAISDDITIAHVFDLLIEGLKYBERV 2977  
QY 294 ----- 293  
Db 2978 RAVNAGPGSPSDATETHVARPKNTPPKIDRNFMWSDIKIKAGNVFEFDVPTGPELPFSKD 3037  
QY 294 -----YSITTAARRVS----- 304  
Db 3038 WTHEGNMIINTDRVKISNFDDRTKIRILSATSDGVYTLT--ARNINGTDRHNKVKVILD 3095  
QY 305 ----- 304  
Db 3096 APSVPEPALRNGDVSKNSIVLRWRPPKDDGGSEITHYVVEKMDNEAMRWVPVGDCTDEI 3155  
QY 305 ----- 304  
Db 3156 RADNLIENHDYSFRVAVNKGQSQPLTTSQPIITAKDPYSHDPKPGQPAATDWGKHFDVL 3215  
QY 305 ----- 304  
Db 3216 EMSTPKRDGGAPISYIIIEKPKQWERAADVLDNCKAHVPELTNGGEYEFVRIAVNR 3275  
QY 305 -----TSTIAYRP-----DSSFMSKI-----MATQLRDLATWY-- 333  
Db 3276 GGPSDPSDPSSTIICKRFLAPFPDKSLNLDITVHAGNGLGWLTPIEASPRPLITWLYNG 3335  
QY 334 ----- 333  
Db 3336 KEIGSNRSGDSLQFQNELTFEIVSVLSRABGRYTLIILKNEHGSFDASAHATVLDPRSPPK 3395

QY 334 -----TTLR 337  
Db 3396 GPLDITKITRDGCHLTWNVPDDGGSPILHYIISKMDLSRSTWSDAGMSTHIVHDVTRLV 3455  
QY 338 YR-----QNPFCPE----- 346  
Db 3456 HRKEYLFRVKAUNAIGSDPLEAVNTIIAKNEFDEPDAPGKLIITDWRDHDIDQWAVPK 3515  
QY 347 SRNRTAVSEFM-----KNT----- 360  
Db 3516 SDGAPISEYIIQKKEKSPYTNVRVPSNKNNTTIPELTEGOEYEFVRIAVNAQGOSE 3575  
QY 361 ----- 360  
Db 3576 PSEPSDMIKRPRVLPKLIITPLNEVRIKCLIFHTDIIHFGEPAPEATWTLSNPLLSN 3635  
QY 361 -----HVLIRN----- 366  
Db 3636 DRSTITSIGHHSVVHTVNCQSDSGSIYHLLLRNSSGIDEGSFELVVLDRPGPPGPMYE 3695  
QY 367 ----- 366  
Db 3696 EITANSVTISWKPPKONGSGSEISSYVIEKRDLTGGGWPAVNVYSAKYNHAVVPRLLEG 3755  
QY 367 ----- 366  
Db 3756 TMYELRVMAENLQGRSDPLTSDQPVVAKSQYTVPGAPCKBELTSDKNHITIKWKQPLSN 3815  
QY 367 ----- 366  
Db 3816 GGSPLIGVDIERRDVNTGRMIKINGQVPPTAEYQDDRVTSNHQYQYRISAVNAAGNGKTS 3875  
QY 367 ----- 366  
Db 3876 EPSAIFNARPLREKPRFYDGLIGKRIKVRAGEPVNLNIPISGAPTPTIEWKRGDKLEE 3935  
QY 367 ----- 366  
Db 3936 GKRISETNSERTLFRIDDSNRDRSGKYVTVAANEFGKDTADIEVIVVDKPSPEGLSY 3995  
QY 367 ----- 366  
Db 3996 TETAPDHI SLHWYSPKDDGGSDITGYIIETFEFGVDDWKVPVGCPTNTFTVKNLVEGKK 4055  
QY 367 ----- 366  
Db 4056 YVFRIRAEITYGASEALEGKPVLVKSPDPPGAPSQPTISAYTPNSANLEWHPPDDCGGK 4115  
QY 367 -----ETPYTI----- 372  
Db 4116 PITGYIVERRERGGEWIKCNYPPTNTSYTVSNLRDGARVEFRVLAVNEAGPHSPKPSD 4175  
QY 373 ----- 372  
Db 4176 PMTAEHQYRDPPEPPKPDRIITRNGVTLSWRPPTDGSRIGKYGYVEMRPNKGDKWKT 4235  
QY 373 ----- 372  
Db 4236 NDIPINSTVTVPSLKEGEYSFRVVAENEVEGRSDPSKPSQPTIEBQPNKPCMELOKVR 4295  
QY 373 ----- 372  
Db 4296 DIVCRAGDDFSIHVPYLAFFKPNAFWYSDNMLDDNNRVHKHLTDDAASVVVKNKSKERDS 4355  
QY 373 ----- 372  
Db 4356 GQYRLQKOTSGFDTATINVRVLDROPSPPTRLRADVFGSDSLTYMNPNDGGSALQNY 4415  
QY 373 -----YG----- 374  
Db 4416 IIEKKEARSWSKVVSFCTVLVFRIRNLVINKEYDFRVAENKYGQSDPANTSEPILAR 4475

```
QY 375 ----- 374
Db 4476 HPDPNTPGPHGIDSTDSITIAWTKPKHGGSPITGYIEKRLLSDDDKWTKAVHALC 4535
QY 375 ----- 379
Db 4536 PDLCKIPNLIAEYEFRAAVNAGQSAVSGSDDLFCRRPPHAPKITSDLSIPDMTV 4595
QY 380 ----- 386
Db 4596 IAGDEFRTVPYHASPRTASWSLNGLEVI PGERIKPDSNDYASWYTKSARDETSYT 4655
QY 387 -MFVENK---TAS----- 395
Db 4656 ITLTNNKGSOTASCHVTVDRLPPQGPLNAYITPDTCTLAWKTLPLDDGGSPITNYVE 4715
QY 396 ----- 395
Db 4716 KLDNSGWWKISSFVRNTHYDVMGLEPHYKYNFRVRAENQYGLSDPLDIIEPMVAKHOPT 4775
QY 396 ----- 401
Db 4776 VPDPGQPKVIDWSGNVTLIWTRPLSDGSRIOGYQIEXRDLNDSSNAYDYIIKGTK 4835
QY 402 ----- 415
Db 4836 YQLYNLINGSYEFRIKAKNAAGLSKSPSPSLRFLKKGFTVPSPPGAPQVRVGKNYVD 4895
QY 416 ----- 415
Db 4896 LWKELPLRDSGRITGYIIERRDIGGAVWVKCNDYNVLDTEYTMNLIEMGDYFRVPAV 4955
QY 416 ----- 415
Db 4956 NSAGRSBSLCTMPIKCEVLGKKPDWITRLQKVAPPGKYDTLQCAASKPSPARWL 5015
QY 416 ----- 415
Db 5016 RKGKEIQWNGGRMTCDSKGVFRLHISNVQTGDDYTCAMNSLGFVNTSGYLKIGSPI 5075
QY 416 ----- 415
Db 5076 INRCPELKLPEGNSKIKIPYSGDQPLTVILKQNEVISDSNDDTHVKVNI PDDYVAIY 5135
QY 416 ----- 415
Db 5136 IANIVKSDGGPYQIEFTNESGSATGEFVYHITGMPAPTGPWGISYINKNSCMLWRPPY 5195
QY 416 ----- 415
Db 5196 YDGLKVSHVYTERKDVSSPHWITVSSCTKDTAFNVQGLIENQEIYIPRVMAVNEGMP 5255
QY 416 ----- 415
Db 5256 LEGLNPRAXDPIIDPPSPAFLKSRSEETLCNPEWKKPESDGAHIQGYWIDKREVGSK 5315
QY 416 ----- 415
Db 5316 HMAGVNATICANQINCINLIEGRQYBRIFAQNVAGLSTAIVRFTSSQDNRTGGFASI 5375
QY 416 ----- 415
Db 5376 DCETATHANCIONHNAQFTCTIPGVKPTISWYKGAREISNGAGITCTQKVTSLPKYVDV 5435
QY 416 ----- 415
Db 5436 FGEDAEYVCRAVNKAGAKSTRATLAINTAPKLVNPPRFRDTAYEDKGENVVIKIPTGL 5495
QY 416 ----- 415
Db 5496 PKPRIHWKGLIENTESGHHYVEKHAVALIIRDSGLSDGPPYRTAENELSGDTAIQ 5555
QY 416 ----- 421
-PLWD-----YL-----
```

RESULT 7

```
Db 5556 VOISDRPPRPFPLIESIGTESLSLWKAQVWDCSDITNYVVERRDDPLSWIRVGNTR 5615
QY 422 ----- 421
Db 5616 FTSMVAGSLTPGKEYDFRIPADNVYGRSDASDTSTLTKTESVKKKPIERKWEIDANGRK 5675
QY 422 ----- 430
Db 5676 LRGKADGPVKDYDSYVFDIYSKFPQPVBEISQQSVYDRYDILBEIGTGAFGVWHRCRBS 5735
QY 431 ----- 430
Db 5736 TGNIFAAKPIPVSHSVEKDLIRREIDIMNQLHOKLINLHDAFEDDDDEMILILFSLSGE 5795
QY 431 ----- 430
Db 5796 LPERITAEGYMTAEVINYMRQICEGIRHMEQNIHLHDIKPENIMCQTRSSNTNVKLID 5855
QY 431 ----- 430
Db 5856 FGLATRLDPNEVKITTTGTAEFGAPEIWRBPGVFTDMWATGVSLLSGLSPFAGDN 5915
QY 431 ----- 430
Db 5916 DDQTLKNVACDWDFALESFKYISEBAKDFIRKLLVRNKEKMTAHECLLHPWLTGDHSA 5975
QY 431 ----- 430
Db 5976 MKQBINRDRYLAVERKURRYEDEFERFLLPTGLRSEYSSLRKLLMEKYKIHDAVDFDRQA 6035
QY 431 ----- 439
6036 APRFVIRPSSQFCYEGOSVKFYCRCIAIATPTLWSHNNIELRQSVKFMKYVGDYVFI 6095
QY 440 ----- 450
6096 INRVKPHDRGEYITRAENHYGSRBEVVFLNVQPLPKQPRYRTESTPVRREPLPYTFWQ 6155
QY 451 ----- 450
6156 ESEETAPSTFLLRPRVWQARDTCKLCLSGKVPVNRWYKQGRELSKYBYAMTHSDGV 6215
QY 451 ----- 450
6216 VTMEIIDCKPSDSGNYCKATNCHGTDETDCWIVEGEWVTPEQAQLAHNFLYSGDRKYI 6275
QY 451 ----- 459
6276 EQPIKAPLPIVTSROYTSSSVQNTSEPOGDKVNSVNSNSGSIKNKKYASNSLQAPGSP 6335
QY 460 ----- 459
6336 SRSRSATKELLPLPDDSLMCKPBTCKPLHDLTHDGEQLILTCVVGDPBPQISWSKNGK 6395
QY 460 ----- 459
6396 SLSSDILDLRYKNGIATLTINEVPPEDEGVICTATNSVGAVETCKKLTIQPLDKNINK 6455
QY 460 ----- 462
6456 RKNVAGDNAPKIVSHLESFRVDRGDVNLACRIITGAHQHFDVWVWLHNKKEIKPKSKDFQYTN 6515
QY 463 ----- 462
6516 EANIYRLOIAEIPPEDGGTYTCEAFNDIGSEFSFCTINVTVPGETKQPSFVKPPTSVSV 6575
QY 463 ----- 465
6576 LEGEGITFECEIDSELNLVWLK 6598
-----WLQ 465
-----
```

T14593  
 syringomycin synthetase - Pseudomonas syringae pv. syringae  
 C:Species: Pseudomonas syringae pv. syringae  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
 C:Accession: T14593  
 R:Guenzi, E.; Galli, G.; Grgurina, I.; Gross, D.C.; Grandi, G.  
 J. Biol. Chem. 273, 32857-32863, 1998  
 A:Title: Characterization of the syringomycin synthetase gene cluster. A link between pr  
 A:Reference number: Z18153; MUID:99047670; PMID:9830033  
 A:Accession: T14593  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-9376 <GUE>  
 A:Cross-references: UNIPROT:O85168; EMBL:AF047828; NID:g3510628; PID:g3510629; PIDN:AA08  
 C:Genetics:  
 C:Gene: syrE  
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F:441-896/Domain: acetate-CoA ligase homology <ACLI>  
 F:914-981/Domain: acyl carrier protein homology <ACP1>  
 F:1529-1984/Domain: acetate-CoA ligase homology <ACLI>  
 F:2002-2069/Domain: acyl carrier protein homology <ACP2>  
 F:2613-3071/Domain: acetate-CoA ligase homology <ACLI>  
 F:3089-3156/Domain: acyl carrier protein homology <ACP3>  
 F:3700-4158/Domain: acetate-CoA ligase homology <ACLI>  
 F:4176-4244/Domain: acyl carrier protein homology <ACP4>  
 F:4768-5224/Domain: acyl carrier protein homology <ACP4>  
 F:5242-5310/Domain: acyl carrier protein homology <ACP5>  
 F:5834-6280/Domain: acetate-CoA ligase homology <ACLI>  
 F:6892-7352/Domain: acyl carrier protein homology <ACP6>  
 F:7370-7437/Domain: acetate-CoA ligase homology <ACP7>  
 F:7790-8440/Domain: acetate-CoA ligase homology <ACP7>  
 F:8458-8526/Domain: acetate-CoA ligase homology <ACP8>  
 F:9015-9083/Domain: acyl carrier protein homology <ACP9>  
 F:946,2034,3121,4208,5274,6330/Binding site: phosphopantetheine (Ser) (covalent) #statue

Query Match 20.4%; Score 500; DB 2; Length 9376;  
 Best Local Similarity 2.5%; Pred. No. 1;  
 Matches 232; Conservative 98; Mismatches 123; Indels 8785; Gaps 80;  
 QY 1 MGRKEW-----VRDVPKM-----14  
 DB 1 MGGQRLSKAPFKLEVTDFEVDGTEQAIAHLSRFRPPESVTGQWESLVRGQPL 60  
 QY 15 -----FVLIS-----ISFL-----23  
 DB 61 HYWFNRLHLVADGICAVLIAHVAVDAYSGLLEGKETLAEGPSYLSFLEEDRAYMASSRY 120  
 QY 24 -----23  
 DB 121 KRDLRFWEQSYAQLPPLLQRRADFAGLTNVPAQSDQVQAMLPRALYNALYQFASCNL 180  
 QY 24 -----L 24  
 DB 181 SVAHVLGVISTYFCRTGVDEIVVGMFVPHNRTNARQKATIGMFSSVIRVGFDPHETL 240  
 QY 25 VSFIN-----29  
 DB 241 VDLMMNIGTQLRRTYRHQFPFAELNLSRLAQNQRHQLFDVLSFSPFGDVTFGSDP 300  
 QY 30 -----29  
 DB 301 ARVLMNDNGYEQTPMAIFVRDYPHSEDIHLDNFNTAYFTFEAQRVQQRIFAMLEAVLN 360  
 QY 30 -----29  
 DB 361 QHDTPLAHFPLMSTSEHQQLATFNDTAHPYPRDVLHQLIEQQAAQRPDACAVRGDSGFL 420  
 QY 30 -----29  
 DB 421 LTYAELNQANQLAHLRIELGVEPDTRVAVSLRGAEMVVALLIGLKAGYVDPDPLP 480  
 QY 30 -----29

DB 481 SARQAYMLEDSSPQAVLITRDLSDNLPSADLPVLVLDGHNDRAQLARQQSVNPDAKALGL 540  
 QY 30 -----29  
 DB 541 QPNHLAVLYTSGTGTGPKGVNHLGVNRLNWARDAYQVNSQDRVLQKTPCGFDLSVM 600  
 QY 30 -----CKVMSKA-----36  
 DB 601 EFPFLLTGAELVMAPPQGHQDDPYLAQVMSDAGITLLHFVPSMLDVLEHRSTRDPQL 660  
 QY 37 -----LV-----38  
 DB 661 RRVLCSEALPRALQRRFEQQLKGVLEHNLVGPTEAIDVTWECRPTDQDGSVPIGRPI 720  
 QY 39 -----NRP-----41  
 DB 721 ANIQMHVLDALGQLQPMGSAVAGELHIGGVARGVYLNQPLSAERFIADPFSNDPQARLYK 780  
 QY 42 -----W-----RGLVLSKIGKY-----53  
 DB 781 TGDVGRMLANGALEYLGNDPQVKIRGLRI-EIGEIEAALAKHPAVHEAVVTAREDIPGD 839  
 QY 54 -----KLDQ-----57  
 DB 840 KRLVAYYTQSAEHTAVDLEALRSHLQOVLPPEYVMPAIYVLEAMPLTNSGKLDKRALPAP 899  
 QY 58 -----LKE-----61  
 DB 900 DGDALISRGVEAPQGEVETLLASINADVLKVEQVRHDFHFFELGHSLLAVKLIERMQV 959  
 QY 62 -----61  
 DB 960 GLSADVHLFGQPTVAALAAAGGSEVVVPANLPIEHCDRITPQMLPINLSQAQIDIV 1019  
 QY 62 -----61  
 DB 1020 VKDMPGGVANQDIYPIAPLQAGILVHHISABQDPYTLKALFALSADRAQLDDFSGALQG 1079  
 QY 62 -----ILRQLE-----67  
 DB 1080 VINRHDILRTAVLWEDLDEPQVWLQRAELHYTELLDPDVGPDVEQLHQRFDHRHRLD 1139  
 QY 68 -----67  
 DB 1140 VRTAPLRIVFSHDPVNDNRWLAMLLCHHLVSDATSLSVILOEQAHLLGQGNALGEAVPY 1199  
 QY 68 -----67  
 DB 1200 RNVQAQRLGVSEAQAFFRDMGLGDIDEPTLPFGLQDVQDQGRNLEBEASVTLAEALNL 1259  
 QY 68 -----67  
 DB 1260 LRAQARQAGSVASIMHLAMARVLGNVSACEQVVGTVLLGRMQAGDADRALGMFINTL 1319  
 QY 68 -----TTIS-----71  
 DB 1320 PLRVDIGATTVAEGIKATHELTALLGHEHAPVLAQRCGSAAPTPLFSALLNYRHSV 1379  
 QY 72 -----71  
 DB 1380 SDVRQKLPGLGSSLRHRRPYQLPVDLNVDDLGDFTLTQAVQOISATRIGEMQVALR 1439  
 QY 72 -----TKYN-----75  
 DB 1440 NLVDALEHTPQAALNSILPDDERELLTGFDNDTAHPYPRDVLHQLIEQQAAQRPDAC 1499  
 QY 76 -----75  
 DB 1500 AVRGDSGTLTYAELNQANQLAHLRIELGVEPDTRVAVSLRGAEMVVALLIGLKAGY 1559  
 QY 76 -----VSKQPVK 82  
 : |||



Db 1560 YVPIDPDLPARQAYMWSAPRAVLTSHELLADLPDLGVPALVLDGRDSDSALLKKQPTG 1619  
QY 83 N-----LTWNT 88  
Db 1620 NPDAKALDQPNHAYLYTSGTGTGPKGVNHEHGLVGNRLWARDAYQVNSQDRVLQKT 1679  
QY 89 EF-----POYILAGPIQYISITYLW 110  
Db 1680 PFGPDVSWNEFFPLLTGABLMARPSGHQDPDY--LAQVSDAGITLLHFVPSMLDVFL 1737  
QY 111-----DFVSTQLRK-----119  
Db 1738 EHRSTEDF--PQRRVLCSEALPRALQRRPFQHLKGVLEHNLNYGPTAAIDVTAWECRP 1795  
QY 120-----119  
Db 1796 TDPGDSVPIGRPIANIOMHVLDALQLOPMGVAGELHIGGIGVARGYNLQNPOLSARFIA 1855  
QY 120-----PA-----121  
Db 1856 DPFSNDPOARLYKTGDVGRWLANGALEYLGRNDFQVKIRGLRIETGEIAALAKHAPVHE 1915  
QY 122-----KYVYSOYHTAK-----133  
Db 1916 AVVTAREDPGDKRLVAYITQTAREHTAVDLETLRGHLQOVLPEYVMPAIYVULLEAMPLTS 1975  
QY 134-----133  
Db 1976 NGKLDKALPAPDGDALISRGYEAEPQGEIEQIAVIWODLLGVEQGRHDHFFELGCHSL 2035  
QY 134-----133  
Db 2036 LAVSLIDRLRKQDLNANVTFTAPSVRMELALISQEKALFRAPANRIPAHCTOLTPTDM 2095  
QY 134-----133  
Db 2096 LPLVELSVAQIELIASVPGGMANTQHIYPLAPLOEGILFHYLLMLERDAYLVRSTIEFD 2155  
QY 134-----TITFRP-----139  
Db 2156 SRARLDAFLEGLQTVIDRHDVLRSSVHVGLPQAVQVHVRQAQSPHITLTLRPDEBALSQ 2215  
QY 140-----139  
Db 2216 LDRISDPGRRLDLRQAPELLAYIARDPDSERWLLALIDHHMISDHVTLLEILBEIRLLM 2275  
QY 140-----PP-----141  
Db 2276 RGSAELLPPQPYREFVAQTASPSSAHEAYFTGRADVDSPFAPFELLEVOGDNVVEE 2335  
QY 142-----CGRVPSMT-----149  
Db 2336 SELALSSDLCAIRIQARERGMSPAVLPHVAQVLAQVLAQVLAQVLAQVLAQVLAQVLAQV 2395  
QY 150-----149  
Db 2396 AERAMGFMFMTLPVRVOLATQSQBELVMATHRDLSELLSHEQASIALAQAQVATGVPL 2455  
QY 150-----CLS-----152  
Db 2456 FSSLLNYRHQNESQLOQWPGRLRLDLSAERTNYPLCLSVNDYGSIDLGLLIHSVQPADPQRL 2515  
QY 153-----EMLNVSKRN-----161  
Db 2516 CAMMQCALEQLTDALHTPQKEVTQDLVPAERNLLLETFTNQTRQDPTDLCIQHLEFA 2575  
QY 162-----161  
Db 2576 QVTRQPDIAVAFOAQRSLYAEELNRQANRLAHHLIGLIGDPDRVAICVERGVEMVWGLL 2635  
QY 162-----161  
Db 2636 GVLKAGAAVPLDPAYPAERLAYMIEDSQPAALMTQRHQBLYLTLTLPLVLLDDQDKT 2695

QY 162-----DT-----163  
Db 2696 FTERDDNPVEALGVNRNLAYVIYTSGTGNGPKGVMIHRGLVNSYVDAARLFDLSPTDTV 2755  
QY 164-----163  
Db 2756 LQONTNLPDLSVEIIFPALLAGATLTPSRIFGSEGTENHGINPTVLHLTAAHWHTLVAE 2815  
QY 164-----163  
Db 2816 WHKQPOVAEQRLQHVRLINVTGDALSQAOKLKLWDEVRAHTRRLINTYGTPTATVSTAA 2875  
QY 164-----GEQCGNFTTFNPMFFNVRWNTKLYV-----190  
Db 2876 VSHDAAAGSEGSNATIGKPM-----NTRIYLLDAHQQPVPYGVAGEIYIGDGVARG 2929  
QY 191-----190  
Db 2930 YLNLEEVNAERFLADPFSESPARMYKTGDLARYMADGRIEYLRNDQVKVRGFRIBLG 2989  
QY 191-----190  
Db 2990 EIEARLGNCTGVKEAVIAREDNPGDKRLVAYVVAQPOSQLTADLRAELAPOLAEYMLP 3049  
QY 191-----GPTKVV-----197  
Db 3050 SAFVLLDELPLTPNKLDRKALPAPDADALISRGYEAPOGETETQIVAIWODLLGIEQVG 3109  
QY 198-----197  
Db 3110 RHDHFFELGHSLLAVSLIERLRKQGLNANVTFTAPTVRMELALISQDKHVLFOVPAN 3169  
QY 198-----197  
Db 3170 RIPACTQLTPTDMLPLVELSVAQIELITSAVAGGANIQTIDYPLAPLOQGLFHYLLNRE 3229  
QY 198-----DSQ-----TIYFLGL-----207  
Db 3230 RDAYLRSMIEFDSRRLDAFLEGLQTVIDRHDILRSSVHVGILPQAVQVHVRQAQLPVH 3289  
QY 208-----TALLRY-----214  
Db 3290 TLTLPBEDALSQDLRLSDPGRRLDLRQAPELLAYIARDPNSERWLLALIDHHMISDHV 3349  
QY 215-----214  
Db 3350 TVELILEIRLLMRGQSADLLPPOPYRDFVAQTASPSSAHEAYFTRRLADVDSPAPFE 3409  
QY 215-----214  
Db 3410 LLEVOGDNVVEAKLALNSDLCIRIQARERGMSPAVLPHVAQVMAQVMAQVMAQVMAQV 3469  
QY 215-----214  
Db 3470 GTVTGRLQGTAGARAMGFMFMTLPVRVOLTQQAQELVMATHRDLSELLSHEQASIAL 3529  
QY 215-----214  
Db 3530 AQRCSSVATGVPLPSSLLNYRHQEDNRLQWPGMRLLDGTERTNYPLCLSVNDYGSSELDL 3589  
QY 215-----AQRN-----218  
Db 3590 IIHSMQANPQRLCAMMQCALEQLTDALHTPQAVTQDLVPAERNLLLETFTNQTRQD 3649  
QY 219-----218  
Db 3650 YPTDLCIQHLEFAQVTRQPDIAVAQVQQLSYADLNRQANRLAHHLISLGLVDPDRVAI 3709  
QY 219-----218  
Db 3710 CVERGVEMIGLLGVKAGAAVPLDPAYPAERLAYMIEDSQPAALLTLPLGLQRLPAL 3769



Db 5930 PACIYTSSTGVPKGVIVTHNGIVRLVQDNQYDFSAEDRVAFSSNPAPDASTPEIWGA 5989  
Qy 324 ----- 323  
Db 5990 LLNGQSVIIEBPVLEPFAAALLKRHGVAMISSTALFNLYAGLIPALAGLRMWC 6049  
Qy 324 -----QIR----- 326  
Db 6050 GERADPASFRVREHSAQVRLFNFGYGTGTCATSYBIFDVLPTLSLPICKFNANVRV 6109  
Qy 327 -----DLA 329  
Db 6110 YVLDARBPVPMGVVDIYIGGTGVALGYLNRPELTAERFSEDPFQOAGARLYRTGDLA 6169  
Qy 330 TWV----- 332  
Db 6170 RWLPDGNLEYLARNDQGVKRGFRVELGEIBSVLHLYDGVRSVVVAHEASPGDTRLVAY 6229  
Qy 333 YT----- 334  
Db 6230 YTVHAGVAAPDEDLRAQLSADLAEMYVPSVFVRDLALPLTLNGKVDOKALFVPLNAMF 6289  
Qy 335 ----- 334  
Db 6290 BRTVAPEGATEQALAEIPOSLLGLERVRHGDGPFELGCHSLAAQLVSRVRQOLNGDMA 6349  
Qy 335 ----- 334  
Db 6350 LRQLFNHPTVAELAKVVDGLQTVDSDSIEPIERNAPLALSFQQRLWFLDRLDPGASSAY 6409  
Qy 335 -----TLR----- 337  
Db 6410 HPMWSLLIRGELDHRAKALDRLVARHESLRTTFELHGEQPVQVIAAADSGFALAEDDL 6469  
Qy 338 -----YRQNP----- 342  
Db 6470 RSQYEQASLNASRIADSEAAAPFDLRQGLIRGLRLRLADDEHMLLITQHHIISDGWSV 6529  
Qy 343 ----- 342  
Db 6530 GVLINEFTALQATEQRPDPLPALSIOYADYAAWQRTFTGERLAEQADLWREHLGAP 6589  
Qy 343 -----FCB----- 345  
Db 6590 TLLSLPTDRPRVQSVYRGGAVPVTIDAALHORLERFCQAHNVTLFMGLLSAWSVLMTRL 6649  
Qy 346 -----PSRN----- 349  
Db 6650 GNERDVIQVPSANRGRTETENLIGFVFNALALRVDLTONPSVAQLLEQVRQTTLAAHEH 6709  
Qy 350 ----- 349  
Db 6710 QDIPFEQVIEALQPPRSMSHSPLCQVALSLDNTSIGGELKLPGLSLHPVLQAHETAQFDL 6769  
Qy 350 -----RTAVSEFMKNTHVLI----- 364  
Db 6770 MLTLASENGALNGVIEASDLFDRSTVERFAQHFTLLLEAMVEDVAQVGLPLLSPAQR 6829  
Qy 365 ----- 364  
Db 6830 LASPALLQPKAVFASGLMVHQRFEQFAAAHPQNTIALVFGREHVSQALNRDANLAHELL 6889  
Qy 365 ----- 364  
Db 6890 AQGVPRDPRVAILAERGTMICAVLAVLKSAAVYVPLDPTPTERLGLVLLTDSAPVALLA 6949  
Qy 365 -----RNETP-----YTIYGT----- 375  
Db 6950 QSACLDVLPASVPLTLGADNDSADLADESVDNRNPREPALGLSREHLYAVIYTSGSGL 7009  
Qy 376 ----- 375  
Db 7010 PKGVLEHGNVARLFDATAGFNFCHEDVMTFFHSFAPDFSWSWEIWCALCYGKLVIPS 7069

Qy 376 -----LDMSSL----- 381  
Db 7070 EVARSPDDFVALVCEQVTVLNOTPSAPROFIQARERSPOSHALREVVGSEALDFRSIQ 7129  
Qy 382 ----- 381  
Db 7130 PWTARTPLSRTRLNVMGITEITVHATYVYPIQSEIDTAMPSLIGPADDLCRLRDDYQ 7189  
Qy 382 -----YNNET----- 386  
Db 7190 QPVPVGNGEIYIGGAGVARHYLNRTELNAERFIADPYALQSGARLYRTGDVAHYRSOGG 7249  
Qy 387 ----- 386  
Db 7250 VVNVGRNDSQIKIRGFRIELGEIEAQLLACPEVREAMVILREDRPGDKRLVAYLIAEDGA 7309  
Qy 387 ----- 386  
Db 7310 APESALLRSQASVLAEHMLPSPFVTLTETWPLTTNGKLDRAALPAPDQFATVSRDYEAPL 7369  
Qy 387 ----- 386  
Db 7370 GAIEITLAAWQELLGVERVGRQDHPFELGCHSFLVISLIERLQIGLGLLDVSTVFSAPT 7429  
Qy 387 ----- 386  
Db 7430 LQAMAAVLGAGTAABRVPANLI PVDCTALT PDMPLV KLTRQBLEHIVADVPGGVANVQ 7489  
Qy 387 ----- 386  
Db 7490 DIYPLSSLOEGILPHLLQSEGDAYLMRTLATPDSRALLDKFLGALQVVINRHDMRSSL 7549  
Qy 387 ----- 386  
Db 7550 RWQGLPOPQVVVHQAOQLPVIQDTPAGDALQMLRERTNTYHMRDLQOAPLIAAYITY 7609  
Qy 387 ----- 386  
Db 7610 DTRQEKWLMALLDHHLSIDNVTLRLIMGEIQAVMDGRADALPPSQPYRNFARAAVCVSA 7669  
Qy 387 ----- 386  
Db 7670 EHEAYFRQLLGDVDTTTPYQVLDVRGGDATILRSVQDLSDDLARIHSTARAQCVPTSV 7729  
Qy 387 -----MFVB----- 390  
Db 7730 LFHAMWGLVVAATSGRDRDDIGFVTLSCRSQGTSGANHALGMFINTLPMRIRLQONSVRDI 7789  
Qy 391 ----- 390  
Db 7790 VQDAYQQLSGLLTHERRAPLALQRCSAVDASLPMFTVLNCRHGDVNTAGENIEDMGEE 7849  
Qy 391 ----- 390  
Db 7850 QEGVHFLGSETRTNYPYIEIAVANEANGFSLTAQSIDGIDPHRIAAYLQGAELVNALEQ 7909  
Qy 391 -----NKTASD----- 396  
Db 7910 DPARLASSLEVIPOAERQOLLNDFNDTASDPAPVPIHALFEDQVRNRPDAVALVYEDRQ 7969  
Qy 397 ----- 396  
Db 7970 LSYRQLNRRANHVARQLQLGVQPDERVAICABRSLDMIAGLLGVLSKGAAYVPIDPAHP 8029  
Qy 397 ----- 396  
Db 8030 ADMRAFMLQDSQPRALLTOGALSPLVGDTPLMLLDSAESLLAADDOAFDANPVVDGLTAE 8089  
Qy 397 -----SNKTT----- 401  
Db 8090 NLAYVIYTSSTGSGKGMVMEHRSVFNFWNLVTRTTHQHCPTPATVALNAGFFPDMSIKG 8149

QY 402 ----- 401  
Db 8150 ISQLFSGHKLVIIPIQLLRANGSELDFLEAHQVHFADTPSQLODTLLSAGLLERSSQVP 8209  
QY 402 -----PT-----SPSMG----- 408  
Db 8210 SVLLGGEAINASTWEKRNCPTRLNYMGPTCTVDATIDLRDLGKESIGRPIANVQ 8269  
QY 409 -----FORTFIDP----- 416  
Db 8270 VHVLDARGEAPLGVAGEIHIGGSGVARGYLNRLDELSAERFIVDFPFSDAANARLYKTGDL 8329  
QY 417 ----- 416  
Db 8330 GRWLADGTLEYMGRNDFQVKVGRFRIELGETENVLLAVPGIREVVVVIARNDSQGDSDSQ 8389  
QY 417 ----- 416  
Db 8390 LVAVVCGESVAAEHLRSELLRHLPEYVPSAFVQLDALPLTANGKLDRLRALPAPQDALA 8449  
QY 417 ----- 416  
Db 8450 SKVYEAPOGETEEAIEIKALLHLDQVGRNDGFLGSGHSLAVQLLSRLRKLKTRIT 8509  
QY 417 -----LW--DYLD----- 422  
Db 8510 LRELFDAPTVRGLASLVNAAAPGEAOSIPRANRSGRLPLFSQORLWFLDHLDAAGRAY 8569  
QY 423 ----- 422  
Db 8570 HLPWALRLTGALDTAALEATLDRLVARHETLRTFELVDGEPVQKTAAPADSRLLPKQL 8629  
QY 423 ----- 422  
Db 8630 RDLGDBERTSTLARLGOENATQLDLTKGPMRLGHLRLVDAEHVLLIILHHIVSDGWSN 8689  
QY 423 ----- 422  
Db 8690 SVLAQEVSAIYAFSOGQKDPLPALPLQYVDVAAWQORSLDGPALQAQIDFWRKLEGAP 8749  
QY 423 -----SL----- 425  
Db 8750 SVLNPLDLRPRPAIQSYTGWVHFSPALSADLRAFSAQGSTFPVLLAGWSMLSHL 8809  
QY 426 -----FLDEIRNFSL----- 435  
Db 8810 SEQTDVVVGTVPANROHPELPMIGFANTLALRVATDRETRLMDLRKLSLTAAVNH 8869  
QY 436 -----RSPTY-----VNLTPP-----EHRRAVNLS 455  
Db 8870 QDLPPQVVSALQPTENVSHSLPQVMLSLDNTFSLQLPDLVELLDSEHHTTQFDLS 8929  
QY 456 TS-----NSLW----- 457  
Db 8930 LSLVDSSESLGGLOYSDDLFTKTTAQAIMQLFARVLENMVSARQSIGQVLDNTPALPR 8989  
QY 458 ----- 461  
Db 8990 SAHTATIAATVEDEQPEALPYEAPEGETEIALANLWELLKLEKVSRRDDPFKLGISLM 9049  
QY 462 ----- 461  
Db 9050 AVOMASRLRKVLGKPIAVRDLFIETPIAGPARTLDGQARPQHSNLPVVRTGSQRPFLF 9109  
QY 462 ----- 461  
Db 9110 VHPLGGEVOYVRDLAAIDPOVPLYGLAASGLAAGETPELYEVPAMAAARYLAIRQVPKG 9169  
QY 462 -----W-----WL 464  
Db 9170 PYRIGGWSAGGLIAYEMARQLQSSGKLEFTGIIDTSARLEQQAPALSGQFLMDWL 9227

RESULT 8

T15789  
hypothetical protein C41A3.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T15789  
R:Bentley, D.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid C41A3.  
A:Reference number: Z18404  
A:Accession: T15789  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-7829 <BEN>  
A:Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA831.  
C:Genetics:  
A:Gene: CESP:C41A3.1  
A:Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/1;  
/2; 7514/3; 7536/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3  
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
P:6547-6616/Domain: acyl carrier protein homology <ACP>  
P:2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted  
Query Match 20.3%; Score 498; DB 2; Length 7829;  
Best Local Similarity 3.0%; Pred. NO. 0.59;  
Matches 223; Conservative 92; Mismatches 140; Indels 6916; Gaps 74;  
QY 1 MG----- 2  
||  
Db 366 MGFVNEWELNRAISSYFGGTNACAIIEKPKPSLVQKESYAENVLFLSAKSHSLK 425  
QY 3 ----- 2  
Db 426 LQIEEYTFQMAQSDSAMEDIYTVNERKTYDFRAAVFGKDNBEIARKLDGDYSLTNLQ 485  
QY 3 ----- 2  
Db 486 ESTFEVEFGSGNEKMLLRLMYEKNETFHTSTVDYKYLKAEKTCGFPPEARLTALFPFKLTLT 545  
QY 3 ----- 2  
Db 546 PLTYNVSRLISSMATPELLVQVNTLPNKLGRGLQIFCLAVAKVITPESAVQLIKGVVA 605  
QY 3 ----- 2  
Db 606 EANLTDLGDIELKSSKIPBIEIHLKSTKKILPIHISGELKETAKPNLMTFIVNGEIL 665  
QY 3 -----RK----- 4  
||  
Db 666 ELDPVRKVKQLICQLFACGDFPAVKFRGRIVKTPYSLKQFWEVQVTAMTNIVDEQT 725  
QY 5 -----EMVRDVPKMP----- 15  
||  
Db 726 NSLSLDAEIESTVRITVQKFLDIEBDDINLLETGAVDLSLTSEMVEAFGTAVNQTMPFDL 785  
QY 16 -----VLISISFL-----VSFNC----- 30  
||  
Db 786 LEAYPTILNIVDFLTLVTVPTVKATTSIHKKTSSELSTSDINVIACDYQFAGVEGEKEL 845  
QY 31 -----K----- 31  
Db 846 WPTLTSRLTTGKISIRKKQCEGDAGLEVGLLKQDISMFONSFFAIKAEAEFLDPQHR 905  
QY 32 VMSKALYN-----RPWRGLV----- 46  
||  
Db 906 LLLNAYNALEKSGLTSTPDADLFLAISAHSEYRALAEKHINELDERLWMGTVHSMVAGR 965  
QY 47 ----- 46  
Db 966 LAVLMGIRGRAMIVDTTCSSVATALEMAVKSIREGRKFAIVATSQLIQSSKWLKSLTLL 1025  
QY 47 ----- 46

Db 1026 DHHSTNSFSDGSGFCRSGVGVIIKTAEGDSAVIKISSAKSHHCGAVMTPVVSISQ 1085  
Qy 47 -----L 47  
Db 1086 LLEEAGSPYVEGHGTATSGADSAESMAYQKLGSELINSSVKAQFGHCEVASGLIQLMKV 1145  
Qy 48 SKIGKY-----K 54  
Db 1146 SSIKGHIIPSVHNILPSEHNRNENIRLPFAEEKQIDRSADVSGFICTKTVTYTER 1205  
Qy 55 LDQLK-----BILROLETT-----69  
Db 1206 VSQLNDNIQNCYLLPVSAKTDGLKKAACLSLIEWIDNSCESLYDISTTLQKQKTNFKWR 1265  
Qy 70 -----ISTKYN-----75  
Db 1266 TAVGSSHADVVLKQLTSEHNTSLTNWHISTSHHSIGCSTPFHNIPEPEDHYSMFCH 1325  
Qy 76 -----75  
Db 1326 RLRPEPHSNTNSIYHLLAVVVALIRVILKHLKLTNSFAVGFSNLIVLAADAPSHYLN 1385  
Qy 76 -----VSKOPVKNLTWNT-----89  
Db 1386 LLHAFANDVVMKRIARDVTIISNVKLLNLNGEPITTARQAVEATIDQKVVRILPET 1445  
Qy 90 -----PQOYIILAG-----98  
Db 1446 TLILSPRAYEFASQLETIQDYKLGKFTQGGQVDFAGTGPILKIDLPEYFPNRKS 1505  
Qy 99 -----98  
Db 1506 FWLPIPOSVPNEKEKPLIPKSYEFLLKQKQHVQNHVVDSKIVLPGATSIRLVHQLN 1565  
Qy 99 -----PIQNV-----104  
Db 1566 GKPTVELSNIDFLNKITPSEAPSVVKBEEQGLEKLVPGETDAISPKTELQNFNPINE 1625  
Qy 105 -----ITY-----107  
Db 1626 RLNAEVHHTDNIYERFANSHLYTNERFQWDSLYKTYMGKEVRFSVMKDLDLIDGLTQA 1685  
Qy 108 -----107  
Db 1686 IVGCYFENTDNSPFVFTIDQLSILNGDISQQLHAVLKYDSSGNFINGDATVYDALG 1745  
Qy 108 -----LMFD-----111  
Db 1746 NIILHISNVTFKRLNGQAPSLTSKTVDSKITYKKVENEDQKRASKMHLVWFBEFNGWT 1805  
Qy 112 -----PYS-----114  
Db 1806 DIDNTGFFDLGLTSIQAVKLRNAIKSNYPNASSTCVDPYSIDLGLSTLNDPQVTE 1865  
Qy 115 -----TQLRPAK-----122  
Db 1866 TSTGEDDIQDLTDEHKPTRLAENPIGVMAAACRLPGGVSPSELWELLKIGKNASSRIP 1925  
Qy 123 -----122  
Db 1926 ATRVPTRTLISGSKYGNPVEGNFITQDVTQDPDPFKISKAELIDPOORLLLECVQ 1985  
Qy 123 -----YVYSQY-----128  
Db 1986 ECLNSGVITSNVGVFVGLMEKEYQDMMESSILAMLGSMVAIVRNVYIFCYGPSV 2045  
Qy 129 -----128  
Db 2046 TIDTACSSSLVALEMAINALLDNCRSKVIVAGVNLILNEKGQGLRTNGKMLSHQGMSLSP 2105  
Qy 129 -----NHTAKTITFRPP-----140  
Db 2106 DSRASGYGRSDGCVVLMLELAKPNFHYMSTTQSVNVNHHGRSVLSLTAPNGVAHROMLTTSV 2165

Qy 141 -----PCG-----143  
Db 2166 INQSPSLAIDYWEAHGTGTPLGDPBIEFNTLSSILQNIIGSVKASLGHCEASAGTCGLLK 2225  
Qy 144 -----RVP-----146  
Db 2226 LFLMLTYOYVPTLIHFHVNLKNDINAGSIRLPIIGEDSELVSAGISSFGVSGTNAIAAFN 2285  
Qy 147 -----146  
Db 2286 DNNKLEPYPIHKYVILPISAKNQISLDNLEKQLSVIPLTDVPICNIASALANNRSHT 2345  
Qy 147 -----146  
Db 2346 IRNALIVSNGIVNSKMGKPHRVAKKDRVHVKLCDLSLLDASLLQYDVINETYTVASLKN 2405  
Qy 147 -----SMTCLSEML-----155  
Db 2406 PQSPAMKFAIKFUTSLSEYIEIVASDGEELLAVLLANGSLUKWENFNKTMIELPIGSLLT 2465  
Qy 156 -----155  
Db 2466 EFADHDLNSTSSISKSYQTPESHNLDSPELMKLMKLYITGYVDVDMATVYSPVEQFI 2525  
Qy 156 -----155  
Db 2526 ALPNYQFNKOTLWPERLEIVDHYLIGTIDEESEDTLILKNQISLRLHPQFKGKPLDVG 2585  
Qy 156 -----NYSKRND-----162  
Db 2586 TMSIATBALKIRWEIPFSIQNLKTELITLTKPAWLETNVTNRNEDDEGFNVSAIDGRL 2645  
Qy 163 -----162  
Db 2646 FSLNASSVEIQNIEVPAVEVQIPDKVYVLYKECPNAVIRHRNMVYVDSRAEQSPRTANI 2705  
Qy 163 -----162  
Db 2706 VLNEIIGFAPTSDMFIEILGVLPSVHYMVQVDDGALWQFQMISQDKRVLNSIYVLKDAK 2765  
Qy 163 -----TG-164  
Db 2766 GLEIPTIRMHKSTLLSSQBSASIVAAKTQMAVRHKVCLAVGDVIBSGLDIDESQLSTGF 2825  
Qy 165 -----EQ-----167  
Db 2826 SELGIDSLATVDLLRLNQKYPPEIELTSDLPDNPSSIIDLIMIEQLLNEKGITEPSEP 2885  
Qy 168 -----CGN-----170  
Db 2886 NTPKTSLRGRKLSIPAVRAQVLAQIEFVENYNSKQEVQAEAPSSSECSNHLESDATV 2945  
Qy 171 -----PT-----172  
Db 2946 DRTEIRKVSIAVFDLATETLSABDLQSKGFTELGMDLSLIVDFVNLNDKYFPDDEITA 3005  
Qy 173 -----172  
Db 3006 SDIFDYPTVDELSDHIVRKXSSSVPPAASEIMEKTMNGISTSDAETHKLENLSQSFMLL 3065  
Qy 173 -----TFNP-----176  
Db 3066 ENQNSINPTLKMIWSNQTIKLVKPSDGNPLFELNANGQEKIQKHFTGPNNIIDLKGF 3125  
Qy 177 -----176  
Db 3126 HEGSTETLYMSLLNLVKSISKLEIQCRFGVQBFGLSISRAFPMKTVAEEKNPLISPAW 3185  
Qy 177 -----176  
Db 3186 YQNQQVSFVDSOSPITGNWMLITGGLSGIGLEIGKFIANNGAENVILISRRQPTAKALRD 3245

Qy 177 -----MFE----- 180  
Db 3246 ILSTELTHIGLARKTIVILKIKNISAKLIVFQSFKLSFSTKVKYISDLFSKVKTFYFNLIQ 3305  
Qy 181 -----VPRWN-----TKLYVG----- 191  
Db 3306 SKCFSKHPHLFEMLEVYLKQWTVHTIAADINDKEKIRELTKLANVGITGIIHSAGVLK 3365  
Qy 192 ----- 191  
Db 3366 DSKIERQKESFNQVFTPKANGPHVLEIEKHFNKYKINFIIMSSPTAACNREGQLNYGV 3425  
Qy 192 ----- 191  
Db 3426 SNAYLEYQVORRRQKSGCAIQGNWIDTGMATDENVRKFLANLGLFLOHNDKALKYLR 3485  
Qy 192 ----- 191  
Db 3486 ACILTKPELIMVANIDMNVILKNRDLPKDLINTGILPFEDFTGKMNESEPLSNGDFEK 3545  
Qy 192 -----PTK-----VNVDSTI----- 202  
Db 3546 VSMNFSVEDEEVELELKEKVSILMCSPTKLKNKNIMDMGLDSKLIVEFLNFINSTFK 3605  
Qy 203 ----- 202  
Db 3606 ISVNLSDAYNHPTLEKLAHIPEQMTIVDHPVNSVKSEBIPKSTDPCTPIFGINIFPDNKN 3665  
Qy 203 ----- 202  
Db 3666 DFDKAKSTAVKLENGEQLPTAGKAVSVVGKSIDRVVSKIKBAAPQIKLCQESSKCVL 3725  
Qy 203 ----- 202  
Db 3726 MLTGQSQVPMGRQLVENYEIFRTTLOSLCKKCDVYLQGDVSLWEILFNTDHYKLLQT 3785  
Qy 203 -----YFL----- 205  
Db 3786 KMQPMIFGAYATQMLSLGIVPDYILGHSGVGLVAGVLAGIMSIEDGLRIVERGA 3845  
Qy 206 -----GLTALL----- 211  
Db 3846 MENIAGLGNLLAVOREIADEVLRKPKSVATINSPKQVVPAGTSVLDAALAFVKGQKQ 3905  
Qy 212 ----- 211  
Db 3906 ATYVNOQYPPHSLNLIQETHLVSLRQCLADIKFSAGRTPLVSNVTGQIINTPSEAYIVKHT 3965  
Qy 212 ----- 211  
Db 3966 VSAKPFVDCVETLOAKGVTVWIDAGSAVALATFVKRIIOPTELSKHRIVQTCESKESVD 4025  
Qy 212 ----- 211  
Db 4026 NLVQACLEQSGLPISWTLTYGCGRNADERLVEFPMTNDI IKNDSEFELLEGHQNGKI 4085  
Qy 212 ----- 211  
Db 4086 VVAGAYQLFKTDQLVKLKAAGMELMKNVKFLKPWYIEDNRQYQIOWNSDMTIELIVNSV 4145  
Qy 212 -----LRY-----AORN----- 218  
Db 4146 IVCSLEVEPQNSVLKLETISENEXPFVHDIFYETLFRNGLOYDSGFRRIESARRSKRCF 4205  
Qy 219 ----- 218  
Db 4206 SQIKSPFAPWPLIDSAMHSITASVVRPRPDYFLPVGWSVTMKDNTSFTLPNLHAQTVI 4265  
Qy 219 ----- 218  
Db 4266 TSETDKIQNVNALLAGDTPICEVRNMTIIVLKTPEVHTRIPNSIETVETPPKSEIEIVG 4325  
Qy 219 ----- 218

Db 4326 FDISLPYNOIISENSENWQHLAKTNTVKQKLHNRSKLQDHARVALLDSADARYWDPEYFGIRP 4385  
Qy 219 ----- 218  
Db 4386 SEAKFIDPQORLLILCSVAKLLDSLLITSLTNTGVFIGCSANEFSHIVAYGYKDPRAEW 4445  
Qy 219 -----CTHSFYL----- 225  
Db 4446 SSGTNSALAGRIAHWLKLGPPVTLDTACSSSFYALSAACDALRTGCEYAIIVGTNVLV 4505  
Qy 226 ----- 225  
Db 4506 MHEMTVDVLQNAKMTVDDFCFAFDVANGYKRSVAVCSMLITKSPNIDSVATITNYATGH 4565  
Qy 226 -----VNAMS----- 230  
Db 4566 NGTSSSLFTPNGLSQLEVQMRATNPLEKILEIQTHCTGTGKLDPIEINALSKLVSSACKI 4625  
Qy 231 -----RNLFRV----- 236  
Db 4626 GSVKSNIGHTEGSSGLVSLCSSLMSFRSKYRVAQLHLKCPNTSIKTNKMICRFIGEDADE 4685  
Qy 237 -----PKYI----- 240  
Db 4686 NNSILINNFGTGSNCVSVLKPNAISEHFVSSEVFPYILLSSHSAKSLQKYVQVLCEFI 4745  
Qy 241 ----- 240  
Db 4746 SNSAKSLHDIMMSLFQKIHVHRQFIIFNPKRIAVTSLDGFVEVRDERLEKLEKHPCSVF 4805  
Qy 241 ----- 240  
Db 4806 LKSGVHFDKDKFORVDLPISIVFNNTLHWALDSYRDEIDRHSQQMSFKNIFEYKVLLE 4865  
Qy 241 ----- 240  
Db 4866 TPMPQQNPISKVVCIGRLDLIPKIEIDEVSSFANGIIIVFHPISNSIFEYLKLVKWSLIS 4925  
Qy 241 -----NGTK----- 244  
Db 4926 RNQNVFIICCFENGTSHTWTGTLRSLASEKMI PYKPVSIDKVDALKLEFNHEYMFEAI 4985  
Qy 245 ----- 244  
Db 4986 FYSKRGYVERLKNVPTFKFAPQYBKCLISGDTGIGSAILNELKPKSSVITRKNIAS 5045  
Qy 245 -----LKNTMR----- 250  
Db 5046 EDGKTFPLSSDITRLDISHKFNVYVFLHAGIVNNSLHENVKRDSDIDEMVSIKLOKAKNLMKC 5105  
Qy 251 ----- 250  
Db 5106 CDETSHFVSSIANVLGSGQSNVAFNSGLVTSFLETSTKSTIIHMGPPWVGMLAQP 5165  
Qy 251 ----- 250  
Db 5166 ERREIVKQIESNGWKLPPNQDAISVYPTQPMETHEQIIVFDGDFDTIVARQPHLOKLISE 5225  
Qy 251 -----KLKRK----- 255  
Db 5226 VVEKTKVKEEIKKSLNFEIFEIVGITDISKLNIPFMDLIGDLSLWENLRLYSLKNFN 5285  
Qy 256 --QAPVKEQFE-----KKAKTKOSTT----- 274  
Db 5286 DLELTVSEMFENATYQKLOTYVETLURKAKHSESLSHDRVSSQVSNKEDDTRVAVIGMSAE 5345  
Qy 275 ----- 274  
Db 5346 FSGSSNIHEYWENLMDGICSTGNKNKYLKPNPFGDNKFFNLTDDEARVLDPOVKRFIOHA 5405  
Qy 275 -----TPYFSY----- 280

Db 5406 YLALENSGYVQKHLCRCVAGAPSPSYGRADDDHDDAMRKLFVNMNMSYLASYCLD 5465  
Qy 281 -----TTSALNVTTNVTYSITAA----- 300  
Db 5466 LKGEAVSVYSACSTALVAVANAVKSIQSGMDYALVGAASIAEVSGALSGDFDQKKTWFS 5525  
Qy 301 -----SEWKN----- 359  
Db 5526 KSGVCRPPDKDSBGIVRSGVGCFLKRYSOALLDNDNVHVIKDFAINNDGHSRSPMA 5585  
Qy 301 -----RRVS----- 304  
Db 5586 PNPAGOLKCMTDVLARFTNKERISFVECHATGTTIGDTIEMNSLRTAYSPKNAIGS 5645  
Qy 305 -----TSTI----- 308  
Db 5646 CKANIGHAYAASGLAALVKCAKMLQTGIIPQVNFSEFRDGMGQFFTVNGKKSITSONSL 5705  
Qy 309 ----- 308  
Db 5706 ISIDSGIGTNVHMVIEFPARSOEHWKISSENLILYDMIPISAKTEYSLDHTSEAIKY 5765  
Qy 309 ----- 308  
Db 5766 LQDTNKIAQCSSTFIHSRVPMDSRTYLSVNNNELLKIRTNKKTWNGKSPKIALFPAP 5825  
Qy 309 -----AYR-----PDSPFKSIMATOL--- 325  
Db 5826 QGIQFTNLPNEYLKNSAYREVEYLCELASSFGIPSLGILYPTKQNFPHLIHATQPAQI 5885  
Qy 326 ----- 325  
Db 5886 AIFVQCMAIFRAIKNVFNPTCLIGHSGEYAAAVISGALKTEBALKLIKESLIGKTEK 5945  
Qy 326 -RDLATWY-----TTLRVQ----- 340  
Db 5946 ARMLAVWNYEKOLPSHVHVSIIIDANTKCVGVGPVETIDNLEKYFINNHKIRYNIETKHGP 6005  
Qy 341 -----NPFC----- 345  
Db 6006 HSKMFHCISKFEFFCESFATKVPILPMISSITGSEIKIFDSKYCTMHLTPNVNLELVVD 6065  
Qy 346 ----- 345  
Db 6066 HIMKLDIDIIVEVGPTGVLSNLLAKRNSKIVVVTGCTGKHPKISLGEICQLWSNGVDI 6125  
Qy 346 ----- 345  
Db 6126 RKLTPKLAVDQVPGYCFDERQFGEKNQKISQKLNRFYKESWKPVSSELEETKPVRF 6185  
Qy 346 ----- 345  
Db 6186 VCDGSLNGISEPLVYLLRKSEDIHKNYFDVELFLKSPRCSTVVFVGMDSNPSVHLSLG 6245  
Qy 346 -----PSR- 348  
Db 6246 LIRCQVLSRIDLYKVENFENIAMSIVIKVLTANGLYFRIDSTGLCKGHPRTTEIPSRL 6305  
Qy 349 ----- 348  
Db 6306 HPVSRGHALVFGANGFTGSIIVFRLLQEMGMNVIPISRASIPSCDITNIKOVNVFKSLGP 6365  
Qy 349 -----NRTAV----- 353  
Db 6366 KKPSVINCVGVTSAKNKTSLSQEIVLSPTFGSVNILKLEPFSIEVDKLVNFSLS 6425  
Qy 354 ----- 353  
Db 6426 SWPLLGNDYASANCFVEALTKQSKYIKQFLTLSLPLEGRMYESSSTSKHNLSQI 6485  
Qy 354 ----- 353  
Db 6486 LMQOELTNVLKTAFAESNMNGIVFVSSVNPADIATKSIAPHRTDENGSEVKAVESLKLPK 6545

Qy 354 ----- 353  
Db 6546 STSCFVIAEWKETLGLSILNDANPNPFFSLGDSLSALQVVMKVQKKTDRIVDVNDLFD 6605  
Qy 354 -----SEWKN----- 359  
Db 6606 NPTLOEFTKPVKNLTTTKFPAGNTWKISYDAIPLTNSQTQMFMLRQIDTTTSKYNLIFKIT 6665  
Qy 360 -----THVLI-----RNETPTIYGTL----- 376  
Db 6666 ISYETKFWERFLKYSLSHLIAYQPSYRTVPKSGNSPYQICSLTESPHDPDKRCNLNNAI 6725  
Qy 377 -----DMSSLY----- 382  
Db 6726 SHEPNHLFEICKSTPLRVRVAECDNSRIHIVNQHHILTDGWSMTVLSDTVSSLYAYR 6785  
Qy 383 ----- 382  
Db 6786 GETSPSPKTKQTIQSQVAMGTSKSGDIKEALEYYQNTVHTIIPYDSEGTGNTSPSYVRISKL 6845  
Qy 383 -----YNETMF-----VENKTASDN----- 398  
Db 6846 IPSKIWKLVGLSKLYNTTWYNLALSVPDCAVRSTGQADILLAVAISGRNADNSLIGY 6905  
Qy 399 -----KT----- 400  
Db 6906 FMNALPKTSLPPEILRELBILNIVLSLEKRSFATIPFYQMVQONRKLMEISLFFNFR 6965  
Qy 401 -----TPTS----- 404  
Db 6966 QKLDYPTVSMFGAKCEIEHLSLNNAFDSFTIDETPSLTIVDFDKSKYLDTTVHMPAN 7025  
Qy 405 -----PSMGFQ-----RTF----- 413  
Db 7026 IFLKLANLRNMNTTIPIRTDPSLTFQKGLFTSWRLFTPEALLLSTNTVSYSDLAEKI 7085  
Qy 414 -----IDPLW----- 418  
Db 7086 ENISKDIQKQIAKATSVREDELVLGDCKNYPALLACVFLGLPYAPIDPTWPEPRQLF 7145  
Qy 419 ----- 418  
Db 7146 VKSKVSFTLENCPSCNLKRNFNSRTQFSIYFTSGSTGVKPKVLMABEQSVSFMSTA 7205  
Qy 419 ----- 418  
Db 7206 SKQCMFRSNIRVLDSVKQVFDVSVSNIGSVLNGGVLLISSEHSTIITDQLQKQYAPLPA 7265  
Qy 419 ----- 418  
Db 7266 AVFNGFTDKTMSRLESJETLTIGGETVSDVVIETAMKKFPLRLRTIYGPPTCISWLTN 7325  
Qy 419 ----- 418  
Db 7326 KCKVSTLNGSALGDSLSNETCTCNNSVRGNVQVKGISLARGYITSAPHGTPPSDIYST 7385  
Qy 419 -DYLDLSL----- 425  
Db 7386 GDIVDSKLSLOYIGRMDSQVKCKGVRINISIEKEILICLGLLOIIVVLSNQMLIAFIV 7445  
Qy 426 ----- 425  
Db 7446 DQSKLLHDSLAVTKLNRTQIPDYFVQINKNPLNSGKDKSLLLQAPENIRKSYKREIV 7505  
Qy 426 ----- 425  
Db 7506 VMKNLEEKVINVPSKILGRNVAPTDKFSIGGNSLNAIQIAHRLABELKIEIKAHEILQ 7565  
Qy 426 -----FLDEIRN----- 432  
Db 7566 SNSLKTFCNTLKNKSVQKIPSKPNVITKLRVPSNKNFIYLVHAIIGGTIIPYPSFLQIFP 7625

QY 433 -----PSLRSP----- 438  
Db 7626 KDLSYLGIEFDLKYFSNDLRELAHFAEIAAHAGNKKRIFVNGHSMGGIMSRREIVABELKI 7685  
QY 439 -----TYV--NLTPPEHR--RAVNLS 455  
Db 7686 WGYDIPFVWLFDSWLVRLTNELDIENIKQFITVYFSGLPDSEHRINRAIKLA 7736

RESULT 9  
Ti6580  
hypothetical protein K07E12.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: Ti6580  
R:Fulton, L.  
submitted to the EMBL Data Library, May 1994  
A:Description: The sequence of C. elegans cosmid K07E12.  
A:Reference number: Z18540  
A:Accession: Ti6580  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-13055 <FUL>  
A:Cross-references: UNIPROT:Q09165; EMBL:U00054; PID:g485141; PIDN:AAA50715  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:K07E12.1  
A:Introns: 46/3; 85/2; 201/1; 278/2; 470/2; 817/1; 927/1; 960/3; 1265/1; 1322/1; 1478/3;  
; 6014/3; 6159/3; 6665/3; 7266/3; 7895/3; 8669/3; 8726/3; 9803/3; 10937/3; 12234/2; 1229

Query Match 19.5%; Score 479; DB 2; Length 13055;  
Best Local Similarity 1.7%; Pred. No. 14;  
Matches 216; Conservative 86; Mismatches 135; Indels 12521; Gaps 71;  
QY 12 PKMFVLISFLAVSF----- 27  
Db 113 PKNATSPKSEFLVKVHCDYGVEMDEVDTMFCQDKKWMTTPMCRGGQLCAADNGGCSH 172  
QY 28 -----INCK----- 31  
Db 173 TCISYNDEKIECKPRGWTLDVEKTCIKPKSLCSLSGCTCNGITETQFACSCGNK 232  
QY 32 ----- 31  
Db 233 QKCLLIAGPRRIYIEPQGYEAPGPNINISCTSVAYPPDIYFKNQKVNTPDQNTL 292  
QY 32 -----VMSKALY----- 38  
Db 293 RASQILIKBIYRNEEFTCVSDNIHGSANRTVSIWVTPGSGAPHLKASAGRTSLTVRWE 352  
QY 39 -----NRP-----WRGLVL----- 47  
Db 353 PPSIINRPIITYTLYTNPOOPVKNMKLEVKSEFTREVAIPDLRPTAYYIRVRANDPL 412  
QY 48 -----SKTG----- 51  
Db 413 GPGKLGNOVQIKLPAVRPYVNIWEGDEIRVPWMTAFEDCNVTRADVPVVLWHLKGR 472  
QY 52 ----- 51  
Db 473 PLNKGSKTOHKMKNGVLESTQSCVAENEAGSKTKKINVTGTPSAPERIRYQIDGDK 532  
QY 60 L-----EILR----- 64  
Db 533 VTLQWEPPIITNGPMAGVDVYTTEDPSLPRDQWKVHHIDDPNARTTTLRLNEKTPYTFV 592  
QY 65 -----QLE----- 67  
Db 593 IVGNRLGPGLPSPAPFATTWLAAPVWQLEPSEEMTKPSNDEMIIECGAGVGPCKI 652  
QY 68 ----- 67  
Db 653 IWLWSGTLIEDGKEEPRVYDPTTDAQDRTRSKLIAQSTRSGVATCQAVNSEGSDKKV 712

QY 68 -----TTISTKYNVKQPKV----- 82  
Db 713 PVKILGPSAPLGIPTPTMHTGFDVAMKPKVTNGRITDYVVVYYSKDPDAPLSDWESKTV 772  
QY 83 -----NLTMNTE----- 89  
Db 773 PADTRNLTVNVDDEDTPVVVKVQARTDDPGIISEAYEVTTGRKQVPLSVLEIADPSVD 832  
QY 90 ----- 89  
Db 833 PSTGETIVEPTQPIHFRCAVDGREMPYSVSWLPINASEGDEVPPIPIHSDDSQPHYN 892  
QY 90 ----- 89  
Db 893 SIQVYSTTATKRILLCOARNPDGTVDRHVFIVNKGSAFONPEVIVDPNRRVITWQPP 952  
QY 90 -----FP-----QYIL----- 96  
Db 953 KYPNGEITSNVYITGDESLPVDQWQVFPVDDVTDPKLVLRGALQPEPFFVKIAAVNP 1012  
QY 97 -----AGPIQNYSI----- 105  
Db 1013 HGEGIHTDPKHFDTVSGAPIDAPTDVLPSVISDNTVNTWSPPTQPLGKTSYTVFAPE 1072  
QY 106 ----- 105  
Db 1073 YDDSDFTWQRIISVDAPDGDHGEVTLPKQFNPNTPYKIRISATNDLSEGPASDPVRFE 1132  
QY 106 ----- 105  
Db 1133 TGSGEIPPTITLDPSTSTYVEPLGAATITCTATGVPQPKVHWIKANGETVDSATLQLYD 1192  
QY 106 ----- 105  
Db 1193 LVKDSATCVAENNAKCTQEAWSIQVTPGTAPNEIVLLPMNQENINWETSPDEVNGOI 1252  
QY 106 -----TYL----- 108  
Db 1253 TNYIIHYGEISEDSEFATWQVTIARDVNHKLANLEPKKYAIRVQAVSDRGPVISA 1312  
QY 109 ----- 108  
Db 1313 PQVITKLPLAQAITNPIIQVHPNNSVTIEFTPPDDPENPGKKVKDFVIQVTTDEPDE 1372  
QY 109 ----- 108  
Db 1373 SVWKELKFTDPDDTDDTTIVSIDGENFNPDTKINTRIARGEIDSQPNPTLFTATGDGI 1432  
QY 109 -----WFDYFSTQL----- 117  
Db 1433 APSQSFNVDTEDGVIKVPAGTDYTIKCVSEGYAPDVRVWDSHGNSQLSDGFLRLIIR 1492  
QY 118 -----RKP-----AK 122  
Db 1493 KTLNAKLAENRGGLKETDLTFVAGPCTAPENIQLTANKPTTISVOQYEPSIENGNI 1552  
QY 123 YV----- 124  
Db 1553 YIIYPTLDDDDPHQLGQVQTKPISDWQNVHMDNDGVEGPRKVDIKDFVSTDTAYAVV 1612  
QY 125 -----YS-QY----- 128  
Db 1613 QAINDDGPYPYSNQYTIIRTSRAREGPPVELRVEPDQORSVAQWKEPVTSDVPPIGVEI 1672  
QY 129 ----- 128  
Db 1673 YVVRGKSVSEDDSGAGLNDWIKISIDDDTKLTHKIQNLLLPDYYVFKRAIYDPGVSF 1732  
QY 129 ----- 128  
Db 1733 SEPCIMKTLPGNAPYIQISTGDNVGEVSTTIQILPGSQMTACNATGILPQVKWIKAG 1792



QY 129 -----NH----- 130  
Db 1793 NYEIDPSRVDAGNHAQFSLOVANITETTFNCVAQNPLGHANWTINVLNLEGLEPNWRD 1852  
QY 131 ---TAKT----- 134  
Db 1853 DPTVTSKDGQIIVFNDELPEYLKPPNEWTIQTDDAEQPKDQWESIPSGGAPLTRVEV 1912  
QY 135 ----- 134  
Db 1913 PNMNGTFFYLVDNPEKGIOTPLVMTPKPPSDIRFGKNDDQIIVDFKPAVASEPIK 1972  
QY 135 -----ITP 137  
Db 1973 EYTSVWSTDPNSVKFTTTPADVTSVVVVGLEPDTBYNVQVAABFYEGELASEPVTV 2032  
QY 138 RPP-----CGR----- 144  
Db 2033 KTFPRDVSCECDHGCAPEWNEAGTMEPKCYCHGGFHLTSGKSCERDEEDDATSQAVLQ 2092  
QY 145 ---VPSMT---CLSEMLNVS----- 158  
Db 2093 VTPPSITTKVAPEELLTSGSGEVDSTPETLSVVVGPDGKPLVLDKKGKPIDSSGKPKVKFD 2152  
QY 159 -----KNDTGE----- 165  
Db 2153 ENGDP IAPEGTKLEKNDNGEWYPLVDRNGKPLPVDENNKPITIVIDKGRVVTETDDGT 2212  
QY 166 ----- 165  
Db 2213 FVTSKGQVEVDDLGRPLDEBGNPKYTNENGQFVISVDGAVEGDBEEQVPIPLYVD 2272  
QY 166 ----- 165  
Db 2273 VDDGKYLDEBNEIPVNEGDGPIDVNGCKLEKNEKGKFKPKESTQETPQTKITIVSP 2332  
QY 166 ----- 165  
Db 2333 DGTPLTDASGSPIGLQGPVPTDASGKPLAKDGSPLPTDNNNGYVILPSSKNSVDSQPT 2392  
QY 166 ----- 165  
Db 2393 DDAGRVYVPVLPDGSPLATDSTGNFVNRHGDIVERDDEGKPMGPDGQLLPTDASGNVYI 2452  
QY 166 ----- 165  
Db 2453 PVTGNDVLPDANGNPIYPVVGPDGTPPLTDASGAIVGPDGQPIPTDSNGKPLSKEGY 2512  
QY 166 ----- 165  
Db 2513 PLPVDNQGNYILLPTEIDAAQSLPTDDAGMPYPIVKPDGTPLATDSTGSPINDNGEIIIE 2572  
QY 166 ----- 165  
Db 2573 KDDGRPFGPGLILPTDASGNVYIPAMLGQPLATDASGNVYLVSTEQVTWKSYPVDD 2632  
QY 166 ----- 165  
Db 2633 SDTIHPIWNPDGTPLATDSTGSYVTEGQIIEKDDGRPLGPDQVDFQNYENQKLSNC 2692  
QY 166 ----- 165  
Db 2693 HPLTVLPTDDSGNYIYPVAESGEETKPTDASGKTVYVVRGPDGTPPLTDASGAVIGPDGE 2752  
QY 166 ----- 165  
Db 2753 VIPTDENGIPLSQDGSPLPTDQNGNYILVLTSETPTKTLPIDESGNVVYPIPKPDGTPLA 2812  
QY 166 ----- 165  
Db 2813 TDSTGSVTEGDTIAKDDGKPLGPDGEVLPTDASGNVYVPTVSDQTLPTDDTGKTV 2872  
QY 166 ----- 165

Db 2873 YPIRGPDGTPLPTDASGPIGPDGEIMPTDENGIPLSKDGTPPLPTDNDGNVYIVPSDBET 2932  
QY 166 ----- 165  
Db 2933 SKELPIDDSGNVYIPIPKPDGTPLATDSTGSVTEGDTIIEKDDGKPLGPDGQLLSTDA 2992  
QY 166 ----- 165  
Db 2993 SGNVYVPPGLDSQILATTDVYKPIYTVIGPDGTALPTDASGAAGPDGTPISTDETGE 3052  
QY 166 ----- 165  
Db 3053 LDKGSLPTDDYGNFVFWVSQELPTDAEVQTPITKPDGTLLATDSSNGVNDNGDIIEK 3112  
QY 166 -----OCGNFT----- 172  
Db 3113 DDEGKPLGPDGEVLPTDGTGNFIYPATTSDEVIPTDSSGKPLYTIRGPDGTPLPTDGTG 3172  
QY 173 ----- 172  
Db 3173 SALGPDGEPISTSSGKPLSKDGSPLPTDNNNGYVLPVPTDSTTKALPTDSSGNVPIBIT 3232  
QY 173 ----- 172  
Db 3233 NPDGTSLATDSTGSVFTDDGQIIEKDDGKPLGPDGAILPTDASGNVYIYPVVGPDQALP 3292  
QY 173 ----- 172  
Db 3293 TDETKTVYVVRGPDGTPLPTDASGAVMPDGEPIPTDANGKPLSKDGSPLPTDASGNVY 3352  
QY 173 ----- 172  
Db 3353 LVPSDEVTAKELPTEDESGTIYVPTRADGTPLATDSTGSVFTDDGQIIEKDDGKPLGPD 3412  
QY 173 ----- 172  
Db 3413 GQVLPTDDSGNYIYPVAVGPDQAPFTDKSGKPLYVRGPDGTSPLTDASGAAGPDGEVI 3472  
QY 173 ----- 172  
Db 3473 PTDENGIPLDKGSPLPTDASGNVYIIVPSGELTWASHPTDDTGNVYIPIPKPDGTLLSTD 3532  
QY 173 ----- 172  
Db 3533 STGSVTEGDTIIEKDDGKPLGNGEALPTDDLGNYIYPIITDSEQTSPTEDVGTSVHL 3592  
QY 173 ----- 172  
Db 3593 VRGPDGTPLPTDASGAIGPDGEVIPTDENGVPILDKDGSPLPTDNNNGYVILVPTKESVTK 3652  
QY 173 ----- 172  
Db 3653 ILPTDDSEAVVHPITRODGTPLSTDSNGFVTDNGEIIIEKDDGRPVGPDGQVLSTDSVG 3712  
QY 173 ----- 172  
Db 3713 NFVYVPTESPNDGKPIHPVLGPDGSPPLTDDSGAVIGPDGEVIPTDASGVPLSKLGLPL 3772  
QY 173 ----- 172  
Db 3773 PTDSGNVYIILSSDVTOKELPTDDTGNVYIPIPKPDGTPLGTDTGTSFVSDDGQIIEKD 3832  
QY 173 ----- 172  
Db 3833 DDGKPLGPDGQVLPTDATGNFIYVPLGPDQALPTDSSGKTVYVVRPDPGNPLPTDASGA 3892  
QY 173 ----- 172  
Db 3893 VIGPGGEPIPTDSSGKPLSADGSPPLTDASGNVYVPSDEVTAKELTDESGTIYVPTVR 3952  
QY 173 ----- 172

Db 3953 ADGTPLATDSTGSGVTDGQTIKDEBKPLGPDQVLPYASGNVYIPVIGPDQALPT 4012  
QY 173 ----- 172  
Db 4013 DESGKTVYVRGPDGTPLPDVSAGAVIGPGEVIPDTSNGIPLSQDGTPLPTDNOQNYIL 4072  
QY 173 -----TF----- 174  
Db 4073 VPTSETATKALPTDESGNVYIPITKADGTPLATDSTGTFVTDGQIIIEKDEBKPLGPDG 4132  
QY 175 ----- 174  
Db 4133 QVLPTDSDSGNYIYVVGPDGQTDDESKTVYVRGSHPTDGTGNVYIPITKPDGTLTADS 4192  
QY 175 ----- 174  
Db 4193 TGSFVTEGQIIIEKDEBKPLGPDGQVLPDTSRGNVYVPEVKSDEQLLPTDHTGKTVPV 4252  
QY 175 ----- 174  
Db 4253 HGPDPGTPLTDDSGAIIIGPGEVIPDENGIPLSKDSQLPTDNNGNVYLVPSDEGATKT 4312  
QY 175 ----- 174  
Db 4313 HPTDTSDAVHPITKPDGTPLATDSTGNFVTENGVDVITKDEBKPLGPNQILPTDASGN 4372  
QY 175 ----- 174  
Db 4373 YIYVIGPDQALPTDESKTVYVRGPDGTPLPDTSAGAVISPDGEVIPDANGIPLDK 4432  
QY 175 ----- 174  
Db 4433 DGSPLPTDASGNVILVPSBQDITKTLPTDSDGNVYIPITKPDGTSLATDSTGSGFTEGGE 4492  
QY 175 ----- 174  
Db 4493 IVERDEBKPLGPDGQVLPDASGNVYIPVVGPDGQVLPDGTGKTVYVVGPDGIPLST 4552  
QY 175 ----- 174  
Db 4553 DASGAVIGDGRPIPTDASGRPLDKDGSFLSDASGNVILVPSDAPTNAGPVPVQHQIT 4612  
QY 175 -----NPMFFNVP----- 182  
Db 4613 RPDGTPLATDSSGHFVTEGVDGVIENDKEGRPIGADQVLPDASNNYIFTDVPTQYAVF 4672  
QY 183 ----- 182  
Db 4673 IPTDVVPIELEAPNCDQVDRVDTLLFVSSHTSAPYLDLTKLIENLLTTTPRDFLPK 4732  
QY 183 ----- 182  
Db 4733 IGTLYSATTEITIDISYDGFKELPDSTNEIREIGGIPDVTNALRTAKMILETSRGDT 4792  
QY 183 -----RWNTKLY----- 189  
Db 4793 LVHLHLSFMRSTSSKVYTERIRALPNTRLIHLNEKQWAEADPNVALLRSLCIPSEVPLP 4852  
QY 190 ----- 189  
Db 4853 SNMPTDASGNLLSIPTDEWVTDGTPDDESGFVYIPITKPDGTPLATDSTGSGFVTEGQII 4912  
QY 190 -----VGP----- 192  
Db 4913 EKNEDGKPLGPDGQVLPDTSNGNVYIPVIGPDQALPTDASGKPIYVRGPDGTPPLPTDA 4972  
QY 193 ----- 192  
Db 4973 SGAVIGDGEPIPTDASGKPLAQDGSPLFVDNEGNYIILPTQQVDTKEYPTDETNVIVP 5032  
QY 193 -----TK----- 194  
Db 5033 ITKPDGTLPTDSTGSGFVTENGDRIEFNEBKPLGPDGEVLATDASGNVYVPGSVVEPTA 5092

QY 195 -----VNVDSQTI----- 202  
Db 5093 EPQEVTHGPDGQVLPDASGKPIYVRGPDGIPLPDTSAGAAIGPDGSETIATDENGIPLS 5152  
QY 203 ----- 202  
Db 5153 KDGSPLPDNTGNVYLVPSDEGATEEKPTQSGESIVHPITKPDGTPLATDSTGSGFVTD 5212  
QY 203 ----- 202  
Db 5213 QVIAKEDGKPIGPDGQVLPDTSNGNVYIPVIGPDQALPTDESGKTVYVRGPDGTPLS 5272  
QY 203 -----YFL----- 205  
Db 5273 TDASGAVIGSDGKPIPTDETGLPLNKDGSPLPTDNDGNYILIIPADESVVKALPTDEAKEV 5332  
QY 206 ----- 205  
Db 5333 YPIVQPDGTPLATDSSSGNFVTSSGDIIDIDDEBKPLGPDGQALPTDSDGNVYIPVIGPDG 5392  
QY 206 -----GLTALLR----- 213  
Db 5393 QALPTDESGKTVYVIRGPDGTPLPDTSAGAVIGDGEPIPTDASGKPLSQDGSPLPTDAS 5452  
QY 214 ----- 213  
Db 5453 GNYILVPSDGEVTKLPTDDVGNVYIPITKPDGTPLATDSTGSGFVTDGQIIIEKDEBK 5512  
QY 214 ----- 213  
Db 5513 LGPDGQVLSDDSGNYIYPAVGPNQGTIPTDDTGRVYVVRGPDGTPLPDTSAGAVIGPD 5572  
QY 214 ----- 213  
Db 5573 GEPIPTDASGKPLSADGSPLPTDNNGNVYIVPTDGTSTKSHPTDSDSGNTIYPPVNEGTP 5632  
QY 214 -----YAQR----- 217  
Db 5633 LSTDLSGNFLTNSGEIVDRDEBKPLGPDGQTLPTDASGNVYVYLQVVEETKPLPTDESG 5692  
QY 218 ----- 217  
Db 5693 NIVVPIPKPDGTPLATDSTGSGFVTEGTVIEKDEBKPVGDQVLPDTSRGNVYIPDVT 5752  
QY 218 ----- 217  
Db 5753 PDGQVQPTDVSGKPVYVRGPDGTPLATDASGAALGPDGKPIPTDSNGVPLSEDSPLPT 5812  
QY 218 ----- 217  
Db 5813 DNQGNVYLVPTSETVTKSMPTDNNRNVYIPTMSDGSLLSTDSGSGFVTEGDKVIEKDE 5872  
QY 218 ----- 217  
Db 5873 GKPLGPDGQVLPDTSAGNVYIPVHGQDGTPLPTDASGAVIGPDGSPPLPTDSDGAVIGPDG 5932  
QY 218 ----- 217  
Db 5933 EVIPTDSNGIPLNKDGLPLSTASGNVYIVSABQPGEEIKEIPITKPDGTLTSDSTGNP 5992  
QY 218 ----- 217  
Db 5993 ITENGRIIERDEBKPIGPDGQILPTDASGNVYVYVIGPDGQGLPTDESGKTVYVRGPD 6052  
QY 218 ----- 217  
Db 6053 GTPLPTDASGAVIGDGEPIPTDASGKPLSQDGSLLPTDNNGNVYVLLPSNEETQGLTDD 6112  
QY 218 ----- 217  
Db 6113 ESNVNIYPIPKPDGTPLATDSTGNFVTNGETIEKDEBKPIGPDGQTLPTDSDSGNYIP 6172

QY 218 ----- 217  
Db 6173 VVPGDQALPTDESGKTIYPVHGPDTPLPTDASGASIGPDGPIPTDSKPLFKOGSP 6232  
QY 218 ----- 217  
Db 6233 LPTDSNGNFIIVPSEKRMDEBLPTDSDSKIIYPIPKPGDTPLASDSTGVFTEDGTIIIEK 6292  
QY 218 ----- 217  
Db 6293 DDDGKPLGPDQVLPPTDASGNVIYPIVGPDKGTQPTDSEKTPYVHGPDTPLPTDASG 6352  
QY 218 -----NCHS-- 222  
Db 6353 AVIGDGEPIPTDASGKPLSADGSLPTDNNKGYLVPADEVTKVLPTDDSGNVVHPIT 6412  
QY 223 ----- 222  
Db 6413 RPDGTPLGTASGSFITDDGOAIEKDDEKPIGPDGQILPIDASGNVIYPIVPGDQALP 6472  
QY 223 ----- 222  
Db 6473 TDESGTVYVVRGPDGTPLPTDASGAVIGLGEPIPTDASGKPLSRNGSLTSSGNYI 6532  
QY 223 -----FYLW-- 231  
Db 6533 FVPTDDEKSKKCDISSLSLIIIFVLVNDGGAQNYDQFKAVVGFSRKVDNSPDIIRL 6592  
QY 232 -----NLFRVP-- 237  
Db 6593 AVL SVGEIAVPLPLGGYQEKHELSLISFEIPIVGTETILSPVQAANOQFTSPRTGI 6652  
QY 238 ----- 237  
Db 6653 SKMVVIFADNEEKSTFFIGATYITVKYGTTPKDIINTLIEACEKGLVEIVDDTKHVIDE 6712  
QY 238 ----- 237  
Db 6713 TVPTISSTPVIVDQSGKPLPTDASGNIDNNKPIVIEGEBPTGPDQKLSKNKKGWVY 6772  
QY 238 ----- 237  
Db 6773 PLVDKFGKPVETDNDKPVITVVDNDGNELSKNDGNWIDLSGNEIDTDELGRPLDSEGN 6832  
QY 238 ----- 237  
Db 6833 PYKFDNHNHVIAPQIEEBEETTAPFIIIDGEPINEDDGVYTDKGNVPTNSEGKPI 6892  
QY 238 ----- 237  
Db 6893 DENGQVLPKNEDEGBFVKPKKEADTTQSTIVSPDGSPLPTDASGAAIGPDGPIPTDSNGRP 6952  
QY 238 -----KYI-- 240  
Db 6953 LAKDGSPLPTDNNGRYVILPSGRYSDTETTESGNVIYPIINPDGTPGTSTGNVITS 7012  
QY 241 ----- 240  
Db 7013 IGDIIERDDEKPIGPDQVLTPTDASGNVIYVVGPDGLILPTDATGKPIYVVRGPDGTP 7072  
QY 241 ----- 240  
Db 7073 LPTDASGAVIGPNGEPIPTDASGKPLSQDGSPLPTDVNGNYIMLPSDEVTSQSPLTDES 7132  
QY 241 ----- 240  
Db 7133 NVIYPIPKDGTPLGTSSSGSFITEDQOIEKDEKPIGPDGQILSTLDASGNVIYPDVG 7192  
QY 241 ----- 240  
Db 7193 PDVQTLPTDGDNI SVPTVEATVEFTSKTPEVIHSITKPDGTPSTSTGEFTVEDGQII 7252  
QY 241 -----NGTKLK-- 246

Db 7253 EKDEGKPIGPDGQVLPPTDASGNVIYPIVIGLGOALPTDKSGKTYVVRGPNGTPLPTDA 7312  
QY 247 -----NTW-- 249  
Db 7313 SGAIVIGLGEPIPTDASGKPLSADGSLPTDAVGNVILVPSDDGVIRTHPTDESGNTIYP 7372  
QY 250 ----- 249  
Db 7373 ITKPDGTPLATDSTGAFVTDGQVIEKDDEKPIGPDGQVLPPTDASGNVIYPTVSSDQV 7432  
QY 250 -----RKLKQ-- 256  
Db 7433 LPTDAKPVIVDQSGKPLPTDASGNIDNNKPIVIEGEBPTGPDQKLSKNKKGWVY 7492  
QY 257 ----- 256  
Db 7493 LVDFKGPVETDNDKPVITVVDNDGNELSKNDGNWIDLSGNEIDTDELGRPLDSEGNP 7552  
QY 257 -----APVKEQFE-- 264  
Db 7553 YKFDNHNHVIAPQIEEBEETTAPFIIIDGEPINEDDGVYTDKGNVPTNSEGKPID 7612  
QY 265 -----KAKKTQST-- 273  
Db 7613 ENGQVLPKNEYGEFVKPKKEADTTQSTIVSPDGSPLPTDASGAAIGPDGPIPTDSSGRPI 7672  
QY 274 ----- 273  
Db 7673 SKDGSPLPTDASGNVILVPSGEGVTDLSPTDEAGNIIYPIPKPGDTLLATDSTGSFVADD 7732  
QY 274 ----- 273  
Db 7733 GOIIEKDDEKPIGPDQVLPPTDASGNVIYPIVIGDQALPTDESGKTYVVRGPDGTPL 7792  
QY 274 ----- 273  
Db 7793 PTDASGAVIGPDGPIPTDPSGKPLSADGSLPTDINGNYVLVPSDESAAKVLPTDES 7852  
QY 274 ----- 273  
Db 7853 VVYPIPKPGTPLGTASGSFVTDGQAIKDDGKPIGPDGQTLPIDDSGNVIYVVGVP 7912  
QY 274 ----- 273  
Db 7913 DGOALPTDESGKTYVVLGPDGIPPTDASGAVIGPDGPIPTDASGKPLSADGSLPTD 7972  
QY 274 ----- 273  
Db 7973 NNGNYVLVPADEVTKVLPTDSDGNVHPIITRPDGTPLGTASGSFVTDGQAIKDDG 8032  
QY 274 ----- 273  
Db 8033 KPIGPDQVLPPTDASGNVIYPIVIGPDQALPTDKSGKTYVVRGPDGTPLSTDASGALIG 8092  
QY 274 ----- 273  
Db 8093 LDGEPIPTDASGKPLSADGSLPTDAVGNVILVPSDDGVIRTHPTDESGNTIYPIITKPDG 8152  
QY 274 ----- 273  
Db 8153 TPLATDSTGAFVTDGQVIEKDDEKPIGPDGQVLPPTDASGNVIYPTVSSDQVLPPTDAE 8212  
QY 274 ----- 273  
Db 8213 KPVIVDQSGKPLPTDASGNVIDNNKPIVIEGEBPTGPDQKLSKNKKGWVYPLVDKFG 8272  
QY 274 ----- 273  
Db 8273 KPVETDNDKPVITVVDNDGNELSKNDGNWIDLSGNEIDTDELGRPLDSEGNPKFDDN 8332  
QY 274 -----TTPYFSY-- 280

Db	8333	GHVIAPOIEBEETTPAIFIIIDGEPINEDDGVYTDKGNVPTNSEKPIDENGQVL	8392
QY	281	-----TTSAAAL-----	286
Db	8393	PKNEDEGFVKPEADTTQSTIVSPDGSPLPTDASGAAIGPDGPIPTDSSGRPIKDGSP	8452
QY	287	-----	286
Db	8453	LPTDASGNYILVPSGEGVTDLSLPTDEAGNIYPTIKPDGTLATDSTGSAVDGQIIEK	8512
QY	287	-----	286
Db	8513	DDEKPIGPDQVLPPTDASGNIYYPVIGPDQALPTDESKTVYVVRGPDGTLPTDASG	8572
QY	287	-----NVT-----	290
Db	8573	AVIGPDGPIPTDPSGKPLSADGSKLPTDINGNYLVPADDEVTKVLTDDSGNVVHPIT	8632
QY	291	-----	290
Db	8633	RPDGTPLGTDSGSFITDQIVKNDGKPIGPDQVLPPTDSSDNIYVPSIGSDEQAMP	8692
QY	291	-----NVTYS-----	295
Db	8693	TDITGSVYPLVSPDGTVIEGPPKAKVPGDGVLPPTDASGHFIGPDGPIPTDYGVTVS	8752
QY	296	ITTAARRVST-----STIAY-----	310
Db	8753	DT-----VTTDPCIPLSNDSTGAFITDGTVIENNEDKPIGPDQVLPPTDAIGNYIYPA	8807
QY	311	-----	310
Db	8808	IGPDQALPTDESGNVYVVRGPDGTLPTDVSAGAVIGPDGPIPTDASGKPLSADGSP	8867
QY	311	-----	310
Db	8868	LPTDNNGNVLPADDEVTKVLTDDSGNVVHPITRPDGTPLGTDSGSFVRDDGQAIK	8927
QY	311	-----	310
Db	8928	DDEKPIGPDQVLPPTDASGNIYYPVIGPDQALPTDESKTVYVVRGPDGTLPTDASG	8987
QY	311	-----	310
Db	8988	AVIGLDEPIPTDASGKPLSAGSPLPTDNNGNVLPADDEVTKVLTDDSGNVVHPIT	9047
QY	311	RPD-----SSFWKS-----	319
Db	9048	RPDGTPLGTDSGSFVRDDGQAIKDEKPIGPDQVLPPTDASGNIYYPVIGPDQALP	9107
QY	320	-----	319
Db	9108	TDESKTVYVVRGPDGTLPTDASGAVIGLDEGPIPTDASGKPLSAGSPLPTDNNGNV	9167
QY	320	-----	319
Db	9168	LVAHEVTTKVLTDDSGNVVHPITRPDGTPLGTDSGSFVTDGQAIKDEKPIGPD	9227
QY	320	-----	319
Db	9228	QVLPPTDASGNIYYPVTSSDQVLPPTDAEKFPVVDQSGKPLPTDASGNIIDNNGKPIVE	9287
QY	320	-----	319
Db	9288	GEBPTGPDQKLSKEKEGWYPLVDKFGKPVETDNDKPVITVVDNGNELSKNDGNW	9347
QY	320	-----IMATQLRD-----	327
Db	9348	IDLSGNEIDTDLGRPLDSEGNPYKFDNGHVVIAPOIEEBEATPAIPIIIDGEPINE	9407
QY	328	-----	327
Db	9408	DDGVYTDKGNVPTNSEKPIDENGQVLPKNEDEGFVKPEADTTQSTIVSPDGSPLPT	9467
QY	328	-----	327
Db	9468	DASGAAIGPDGPIPTDSNGRPLAKDGSPLPTDNNGRVILPSPGRYSGDTTETDESIGNVI	9527
QY	328	-----	327
Db	9528	YPIINPDCTPLGTSTGNYITSIGDIIRDEDEKPIGPDQVLTDDASGNIYYPVVGPDG	9587
QY	328	-----	327
Db	9588	LILPTDATGKIYVVRGPDGTLPTDASGAVIGNGEPIPTDASGKPLSQDGSPLPTDVN	9647
QY	328	-----	327
Db	9648	GNVIMPLSDEVTSOSLPTDESGNVYIPTKPDGTPLGTSSGSPITDGGQIIEKDEGKP	9707
QY	328	-----	327
Db	9708	IGPDQILSTDASGNIYYPDVDPVQTLPTDGMISVPTVEATVFTSDKTPEVIHSITK	9767
QY	328	-----	327
Db	9768	PDGTPLSTDSGEFVTEDEGQIIEKDEGKPIGPDQVLPPTDASGNIYYPVIGLQALPT	9827
QY	328	-----	327
Db	9828	DKSGKTVYVVRGNGTPLPTDASGAVIGLDEGPIPTDASGKPLSADGSPPLPTDAVGNYL	9887
QY	328	-----LAT-----	330
Db	9888	VPSDDGVIRHTPDSEGNVIYPTKPDGTPLATDSTGAFVTDGQVIEKDEGKPIGPDG	9947
QY	331	-----	330
Db	9948	QVLPPTDASGNIYYPVTSSDQVLPPTDAEKFPVVDQSGKPLPTDASGNIIDNNGKPIVIEG	10007
QY	331	-----WVY-----	333
Db	10008	EEPTGPDQKLSKNGKGEWYPLVDKFGKPVETDNDKPVITVVDNGNELSKNDGNWI	10067
QY	334	-----TTLRYRQNPFCBPSPNRRTAVSFMKNTHVLI-----	364
Db	10068	DLSGNEIDTDLGRPLDSEGNFY-----KFDNGHVVIAPOIEEBEATPAIPF	10116
QY	365	-----	364
Db	10117	IIIDGEPINEDGVYTDKGNVPTNSEKPIDENGQVLPKNEDEGFVKPEADTTQSTI	10176
QY	365	-----	364
Db	10177	VSPDGSPLPTDASGAAIGPDGPIPTDSSGRPIKDGSPPLPTDASGNIYLVPSGEGVTD	10236
QY	365	-----	364
Db	10237	LPTDEAGNIYPTIKPDGTLLATDSTGSAVDGQIIEKDEKPIGPDQVLPPTDASGN	10296
QY	365	-----	364
Db	10297	YIYVIGPDQALPTDESGKTVVVRGPDGTLPTDASGAVIGPDGPIPTDPSGKPLSA	10356
QY	365	-----	364
Db	10357	DGSPLPTDINGNYLVPSDESAKVLPTDESGSVVYPTIKPDGTPLGTSSGSIYTEDGQ	10416
QY	365	-----	364
Db	10417	LVGKDEGKPVGPQVLPPTDSAGHYVYITGADROLTTDAAGKPIYSVFNEDEGIQLP	10476
QY	365	-----	364
Db	10477	DSSGVAIGHDELVPTESTNGVPLNKDGTPLPTNDSGHFVLVLPATVNDKPTDEVIVS	10536

QY 365 ----- 364  
Db 10537 INPDGTLTGTDSTGAFVTEDEGPIIENDEDEKVPDGOVLPTDSDSGNIYIPVIGPDQA 10596  
QY 365 ----- 364  
Db 10597 LPTDESGKTVPIRGPDGTPLTDASGASIGPDGEPIPTDASGKPLSKDGSPLPTDNDGH 10656  
QY 365 ----- 364  
Db 10657 YVLVPDDSTIKAFPTDESGNAVYPIITRPGTPLGTDSGSFVTDGTIIENDEDEKPIG 10716  
QY 365 ----- 364  
Db 10717 PDGOVLPTDASGNIYIPVIGPDQALPTDESGKTVYPVHGPDGTPLTDASGAIGPDGE 10776  
QY 365 ----- 364  
Db 10777 PIPTDASGKPLSQDGSALPTDNNGNFILVPSDKSTTKTLPTDESGNIYIPITKPGCVLFA 10836  
QY 365 ----- 364  
Db 10837 TDSTGNYVTDEGELIEKDDNGYPLGDKRVLPTDGSNGVIYPVAGSDEKILPTDNLKVV 10896  
QY 365 ----- 367  
Db 10897 YPIITRPGSPLATDSTGVFTGDTIVERNDEGKPIGPDGOVLTTDSDSGNIYIPVIGPDG 10956  
QY 368 ----- 367  
Db 10957 EPLGTDASGKTVYVRGPDGTPLATDAFNAVIGPDGEPIPTDASGKPLDQSGFPLPTDNN 11016  
QY 368 ----- 369  
Db 11017 GNYILVPSDEALGKILPTDENGNNVSVTNPDGTPLATDSTGSFIASNGLIVERKDEKCP 11076  
QY 370 ----- 369  
Db 11077 IGPDGOVLPTDASGNIYIPVIGPDQALPTDESGKPIYVPFTEGDTQPLTSDTGFAIGPD 11136  
QY 370 ----- 373  
Db 11137 GELVPTDSANGVPLSKDGSPLPTDASGNIYILPDSGVTTANPTDENGVAIYIPITKPDGTLL 11196  
QY 374 ----- 373  
Db 11197 ATDSTGSYITQGGQLIEKDNTEGKPIGPDGOVLPTDGSNGVYYPVVGPDQALPTDDTGNV 11256  
QY 374 ----- 373  
Db 11257 VYPVINADGSLATDSSGSFITENGKIVAKDEGKPISPDGOVLPTDASGNIYIPALGPD 11316  
QY 374 ----- 373  
Db 11317 GSILPTDSNGKSIYVVRGPDGTPLTDBEFAIGPDGKPIPTDTSKGKPLSADGSPLPTDN 11376  
QY 374 ----- 373  
Db 11377 NGNYILVLSGVTEHAPTDENGNIYIPVTPNPDGTPLGTDSGAFITQDGTVVKDEDEKCP 11436  
QY 374 ----- 373  
Db 11437 IGPDGOVLPTDSDSGNIYIPVIGPDGOVLPTDASGKTVHVSVPDGTQPLTDASGAIGPD 11496  
QY 374 ----- 373  
Db 11497 GELVPTDSGRPLSQDGSPLPTDNNGNVALVVSDEATTKVLPTDEGGNIYIHTKPDGSL 11556  
QY 374 ----- 373  
Db 11557 LGTDASGDFITDHGKAVQKDEGKPIGPDGSVLPTDTSNGNIYIPITGPDGNVLPTDSNGK 11616  
QY 374 ----- 373

Db 11617 PVYVFNEDGTQPLPTDSTGSAIDQDQELVSTSTSGVPLAKDGSPLPTNSAGNYVLVSSG 11676  
QY 374 ----- 376  
Db 11677 KSQPTDEHGNVIYPIITKPDGTLLATDSTGSYLTEGQQLVEIDDSCKPLSGDGOVLPIDAS 11736  
QY 377 ----- 383  
Db 11737 GNYIYPALGPDGOALPTDDAGNLVYPIYVYVPGTPLATESTGNYVTENGVEVVGKNTDGKPI 11796  
QY 384 ----- 383  
Db 11797 SPDGOVLPTDASGNIYIPAVGPDGOVLPTDASGKLIYVPHPDGTQPLTDASGYAVAPDG 11856  
QY 384 ----- 383  
Db 11857 SLIPTFSGKPLGKDGSLVPTDNSGRYVLVHDDREVTOIPTDESGNTIYPIITRPGDTLL 11916  
QY 384 ----- 383  
Db 11917 STDSTGIYLTDEGNVIDRDNKGLPGDGOVLPTDGYGNFVVPADSDIGGAKLLPTDEYG 11976  
QY 384 ----- 383  
Db 11977 HTLYPVIRPDGSLSTESSGSFVTDGTVVSKSDGKPLGPAGQVLPTDASGNIYPSIG 12036  
QY 384 ----- 383  
Db 12037 PDGSPLPTDINGKPAYTVIGRYGDLVPTDSLGRAVINIDGSVVPDDEGLPIDOYGVVLPT 12096  
QY 384 ----- 383  
Db 12097 DTTRKCLTVLPTRRPSSFCYVTSIDLILLVIDSSNNIKVLDYRVNMKELIKNLFTEHFNLR 12156  
QY 384 ----- 383  
Db 12157 KHQVRVGLVKYGDGAEIPVSLGDYDNEDDLVRHISESRRLKRAQLGAGLREALDELSIS 12216  
QY 384 ----- 401  
Db 12217 GVDGVPQVILVKNKASDDYSSAVKSLKAERNVTVFVVDAGDDDESQQNSLFEEDTKI 12276  
QY 402 ----- 424  
Db 402 PTPSMGFORFIDPLMDYL ----- DSL 424  
QY 12277 VISQWKGADSSVLGPIADYICKIIVPNVETSRTPRTKATTGTSGRSCSSIDVESDI 12336  
QY 425 ----- 433  
Db 425 LFLDEIRNF ----- 433  
QY 12337 IVLDSSENFTPDEFVSMKDAVASIVDTGFDLAPDVSKIGFVIYSDKVAVPVALGHYEDKI 12396  
QY 434 ----- 433  
Db 12397 ELLEKIIDAEKINDGVAIALYGLNARQQFOLHGRNATKVILITNGKNGNAAAAAED 12456  
QY 434 ----- 433  
Db 12457 LRDYGVQLPAVAVGSNPEELATIKLVGNSNTENVIEVAQSTEIDDDAAALLKAVCGNT 12516  
QY 434 ----- 433  
Db 12517 SPKNSMPAHLTKRDVLAKQFTTAPMLRTTRAVAGGLCNDGIRRPVHFNILVDITSRAS 12576  
QY 434 ----- 433  
Db 12577 ADEPRVLDHLINFPNDRMRDEQHMITINIITVNSKVQNILSNLRADQSLSEQLNAITQQ 12636  
QY 434 ----- 433  
Db 12637 SDDTVSPKLGAGIDALAELSKENYINGAIKMLIVGSDGTSSDDALPAEYANSDFQHNI 12696  
QY 434 ----- 438  
Db 434 ---SLRSP----- 438

Db 12697 IAVSVKRPATDLSKIAGLPTRVVHLDQMSAPNELFDSWIAIYITCDYATASTTRKSTTEK 12756  
QY 439 ----- 438  
Db 12757 MTTLRPYDRKASKEDATNIELIPLSPSSLSVSWTCCTNNKSNYILYTHDTSITKEKWIR 12816  
QY 439 -----TYVNLTPPEHRAV----- 452  
Db 12817 KEVTCRDSFGTHNELPDSHTYTCVMTNERVDNSTALDKNCDSLHDQNTAPEDYV 12876  
QY 453 ----- 452  
Db 12877 KPSPSSNCQCSEKAVLRATCEBMVIDTNRPIATLPATVDECPCKVKAHGRCPKGYIA 12936  
QY 453 -----NLSTNSL----- 460  
Db 12937 KDQCQYDIDCEETNNQCSEGCNTPGSYACACPHGMWRDPLDPFNCVNTANSFKIAAL 12996  
QY 461 -----WWWL 464  
Db 12997 LANYLEANTKNSGSEVTSEKSDGRVNYKATIKSADDKTITFEWSHVPEVVERAFKWL 13054

RESULT 10  
T27935  
hypothetical protein ZK617.1b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T27935; T28031  
R:White, S.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: Z20442  
A:Accession: T27935  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-7160 <W12>  
A:Cross-references: UNIPROT:Q23551; EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617.1b  
A:Experimental source: clone ZK617  
R:Harris, B.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: Z20458  
A:Accession: T28031  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-7160 <W12>  
A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b  
A:Experimental source: clone ZK629  
C:Genetics:  
A:Gene: CESP:ZK617.1b  
A:Map position: 4  
A:Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 593/67/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3  
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 19.3%; Score 473; DB 2; Length 7160;  
Best Local Similarity 3.1%; Pred. No. 2.2;  
Matches 217; Conservative 91; Mismatches 130; Indels 6585; Gaps 70;

QY 1 MGRKEMVRDVPKMFVLISIFLLVSFINKVMSKAL----- 37  
Db 106 VGKPRIIPKGGALIVM-----ECKVKSASTPVAKWMKGVPLSMGGLYHAIFS 154  
QY 38 ----- 37  
Db 155 DLGQTYLCQLEIRGPSSDAGQVRCNIRDQGETNANLALNFEPPDPSRQBRKRSTAS 214  
QY 38 ----- 37  
Db 215 PRPSSRGPGRPSSPKMSKREKTPKTLKPRGSPSKLRSRTSTPYNNEVSQSESRR 274  
QY 38 ----- 37  
Db 275 SSRTDKMEVDQVSGASKRKPDGLPPPGDEKBLRAGSPSTRKSPRSKASPTPSRKSSA 334

QY 38 -----YNRP-----WR 43  
Db 335 GGAASGTTGASASATSATSGSASSDASRDKYTRPPIVLEASRQTCRIGGSVVLEVMQ 394  
QY 44 -----GLVLSKIGKYK----- 54  
Db 395 CHSSTIIEWYRDDGTLVRNSSEYSQSFNGSIKQLVKNLITEKSLGYKCHAKCDYEGOSS 454  
QY 55 ----- 54  
Db 455 AMWKIEQSDVEBELMKHKADEYQKEQKSQTLQAEKTKVARRSKSKSPAPQAKK 514  
QY 55 -----LDQL-----KLEILRL----- 66  
Db 515 STTSRGRQEAASEVEHKRSSVRDPDESQLDEIPSSGLTIPERRRELLGQVGSDEDE 574  
QY 67 -----ETTISTK----- 73  
Db 575 VSEISSELPSFAGKPRKTDIFSCVLLYLQSPPKQDMFSDRTLRLKRTTITTKNESSTV 634  
QY 74 ----- 73  
Db 635 EEKTKLRKTVKVDGELDFKAMVKLKKVKEEGTTEKSGFPLDHDADSTSSVLSQESRSR 694  
QY 74 ----- 73  
Db 695 RGSNAPPAKGLPEQANPPAQLKVKVKGAGGLEKSDSMASLKKLDLKKGIIDNSDCAF 754  
QY 74 -----YVSKQPVKNLTWN-----TEF----- 90  
Db 755 KVQLKKVVKKEVKESTISVKEKNGTSEGIKTEFMEKREKRTTLQKYEKTDGSKKEDKP 814  
QY 91 -----POYY----- 94  
Db 815 KKVSIAPVSTNKSDDEPSTPRHHEKVEEKSTSEELKAKVAGQVQKRNQAQKPEEPKN 874  
QY 95 ----- 94  
Db 875 LLSQIQLKKVTKKAHDDTNELEGIKLVKTVTPKHVADDDSOSESRRSGSVFGEJLRGS 934  
QY 95 -----ILAGPIONYSI 105  
Db 935 RAPDSADNSRDSIRRSIDMRRESVQIELEKSTPLVPSGASGAPKIVEP-ENVTV 993  
QY 106 -----TYLWF----- 110  
Db 994 VENETAILTKVSGSPAPTFRMFKGSREVISGGRFKHITDGKEHTVALALLKCRSQDEGP 1053  
QY 111 -----DF----- 112  
Db 1054 YTLTIENVHGTDSADVKLIVTSDNGLDFRAMLKHRESQAFQKDGEGGGGGGGEKKPM 1113  
QY 113 -----YS----- 114  
Db 1114 TEARRQSIFPGKVEKWDIPLPEKTVQOQVDKI CEWKCTYSRPNKIRWYKDRKEIFSG 1173  
QY 115 -----TQLRKPAY----- 123  
Db 1174 GLKYKIVIEKVNCTLIINNPEVDDTGKTCCEANGVPTHAQLTVLBPPMKYSFLNPLPNTQ 1233  
QY 124 ----- 123  
Db 1234 EYRTKQAVLTCKVNTPRAPLVYVRGSKAIQBGDPRFIEKDAVGRCTLTIKEVEEDQA 1293  
QY 124 -----VYSQY----- 128  
Db 1294 EWTARITQDVFSKQVYVEEPRHTFVVPKMSQKNESDLATILETDVNDKDAEVVWHDGK 1353  
QY 129 ----- 128  
Db 1354 RIDIDGVKFVKNRERLLIINGARIEDHGEYKCTTKDRTMAQLIYDAKKNFIVALKD 1413

Qy	129	----	-NHTAK-	133
Db	1414	TEVIEKDDVTLMCQTKDTKTPGIWFRNGKQISSMPGGKFETQSRNGTHTLKIGKIEMNEA	1473	
Qy	134	----	----	133
Db	1474	DVYBIDQAGLRGSCNVTVLEABKRPILNWKPKKIEAKAGBPCVVKVPFQIKGTRGDPKA	1533	
Qy	134	----	----	133
Db	1534	QILKNGKPIDEMERKLEVEIHKDDVAEIVFKNPQLADTGKWALSGNSAGCTALAPPELV	1593	
Qy	134	----	----	133
Db	1594	KDKPKPPKGPLETNNVTAEGLDLVWGTPDPODEGAPVKAYIEMOGRSGNWKVGETKGT	1653	
Qy	134	----	-----T 134	134
Db	1654	DFKVKDLKEHGEYKFRKALNECGLSDPLTGESVLAKNPYGVPGKPKNMDAIDVDKHCHT	1713	
Qy	135	ITFRPP : :	----	140
Db	1714	LAWPEPDEGGAPITGYIIBERREKXDHQVGQTKDCCBELTDKKVVEDKEYLYRVKAV	1773	
Qy	141	-----PC- 	-----142	142
Db	1774	NKAGFGDPCDHGKPIKMAKKASPEFTGGGIKDLRLKVGETIKYDVPISGEPLPECLWV	1833	
Qy	143	-----GRV- 	-----145	145
Db	1834	NGKPLKAVGRVKMSERGHIMKIENAVRADSGKFTITLKNSSGSCDSTATVTVVGRPTP	1893	
Qy	146	-----    : : -----PSMTCL-	151	151
Db	1894	PKGPLDIADVCA DGATLSWNPDDDDGGDPLTGYIEAQDMNKGKYEIVGKVDPNTTLK	1953	1953
Qy	152	-----SMLNVSK-    : : -----	159	159
Db	1954	VNGLRNKGYKFRKAVNNEGESPLSADQYTIKDPWDSBPKGRPEITDFDADRIDIA	2013	2013
Qy	160	-----	-----159	159
Db	2014	WEPHPKGGAPIEBYIEVVRDPTKEWKVKVRPDTNASISGLKBGKEYQFRVRVANKAG	2073	2073
Qy	160	-----	-----159	159
Db	2074	PGQSBPSEKOLAKPKFIPAWLKHDLNLSITVKAGATVRWEVKIGGEPIDEVKWFKNQ	2133	2133
Qy	160	-----RNDTGE- : :	-----165	165
Db	2134	LENGIQLTIDTRKNEHTILCIPSAMRSDVGEYRLTVKNSHGADBEKANLTVLDRPSKNG	2193	2193
Qy	166	-----	-----165	165
Db	2194	PLEVSDVFEDNLNLSWKPPDDGGEPIEYVEVEKLDATGRWVPCAKVKOTKAHIDGLKX	2253	2253
Qy	166	-----	-----165	165
Db	2254	GQTYQFRVKAVNKEGASDALSTDKDTAKNPYDEPGKTGTDPDVVDADRVSLWEWEPKS	2313	2313
Qy	166	-----QCGCNTF-    : : -----	-----172	172
Db	2314	DGGAPITQYVIEKKGHRDQWECGKVGSGDQTNAEILGLKEGEBYQFRKAVNKA GPGEA	2373	2373
Qy	173	-----TFNPMFFNVP- : :	-----182	182
Db	2374	SDPSRKVAKPRNLKPWIDREAMKTIITIKGVNDVEFDPVRGEBPPPKCEWIFNEKPVDDQ	2433	2433
Qy	183	-----	-----182	182
Db	2434	KIRIESEDYKTRFVLRGATRKHAGLYLTATNASSGSKHSEVIVLGPSSPLGPLEVSN	2493	2493
Qy	183	-----	-----182	182

2494	Db	2494	VYEDRADLEWKVPEDDGGAPIDHYEIEKMDLATGRWVPCGRSETTKTTVPNLQPGHEYKF	2555
183	Qy	183	-----	182
2554	Db	2554	RVRVKNKEGSDPLTTNTAILAKNPYEVPGVKDPELVWDKDHVDLAWNAPDDGGAPIE	2613
183	Qy	183	-----RWNTKLYV-----	190
2614	Db	2614	AFVIEKKDKNGRWEALVVPDQKTATVPNLKEGEEYQFISARNKAGTGDPSDPSDRVV	2673
191	Qy	191	-----	190
2674	Db	2674	AKPRNLAPRIHREDLSDTTVKVGATLKFIVHIDGEPAPDVWSPNGKIGESKAQIENEP	2733
191	Qy	191	-----	190
2734	Db	2734	YISRFPALKALRKSGKYTITATNINGTDSVTINIKVSKPTKPGPIEVDTVFEDRATL	2793
191	Qy	191	-----GPTKVNDS-----	199
2794	Db	2794	DMKPEDDGGEPFIEFYIEIKNTKOGIWVPCGRSGDHTFTVDSLNLKGDKHYKFRVKAUNSE	2853
200	Qy	200	-----	199
2854	Db	2854	GPSDPLETETDILAKNPFDRDRGRPEPTDWDSDHVDLKWDPPLSDGGAPIEEYQIEKR	2913
200	Qy	200	-----QTYIFLGITA-----	209
2914	Db	2914	TKYGRWEPAITVPGGQT-----TATVPOLTPNEEYEFVRVAVNKGKGGPSDPSDASKAVIA	2967
210	Qy	210	-----	209
2968	Db	2968	KPRNLKPHIIDRDALKNLTIKAGOSISFDVPVSGEPATVTWHPDNREIRNGGRVKLDNP	3027
210	Qy	210	-----LLRLRYAQRNCTHSF-----	223
3028	Db	3028	EYQSKLVVKMERGDSGTFTIKAVNANGEDBATVKNVIDKPTSPNGPLDVSVDHGDHVT	3087
224	Qy	224	-----	223
3088	Db	3088	LNWRAPDDGGIPIENTYIEKYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVAAV	3147
224	Qy	224	-----	223
3148	Db	3148	NAEGSDPLETFGTTLAKDPFDKPGKTNAPETDWDKDHVDLEWKPPANDGGAPIEEYV	3207
224	Qy	224	-----	223
3208	Db	3208	EMKDESPFVNDVAHPAGQTNATVGNLKEGSKYEFIRAKNKAGLGDPSDSASAVAKAR	3267
224	Qy	224	-----	223
3268	Db	3268	NVPPVIDRNSIQBIKVKAGQFSLNIPVSGEPTTITWTFEGTPVSDRMLKNNBDGKT	3327
224	Qy	224	-----YLVNAMSRY-----	232
3328	Db	3328	KFHVKRALRSDTGTYYIIKAENENGTDATVETVTVLDHPSSPRGPLDVTNIVKDGCDLANWK	3387
233	Qy	233	-----	232
3388	Db	3388	EPEDDGGABISHYVIEKQDAATGRWTACGESKOTNFHVDDLTOGHEYKFRVKAUNRHGDS	3447
233	Qy	233	-----	232
3448	Db	3448	DPLEAREALIAKDFDRADKPGTPPEIVMDKHADLKWTPPADDDGGAPIEGYLVEMRTPS	3507
233	Qy	233	-----LFRV-----PKYI-----	240
3508	Db	3508	GDWVPATVCGELTATVDGLKPGQTYQFVRVKALNKAGESTPSDPSRTWAKPRHLAPKI	3567
241	Qy	241	-----NGTKL---K 246	

Db 3568 NRDMFVAQRVAGOTLNFVNVGEPAKIEWFLNGSPLSGGNTHIDNNTDNNTKLTTK 3627  
QY 247 NTMR----- 250  
:|  
Db 3628 STARADSGKYIVATNSGKDEHEVDNILDIPGAPEGPLRHKDIITKESYVLKWDPLDD 3687  
QY 251 -----KLK----- 253  
:|  
Db 3688 GGSPIITNVVKEQBDGGRWPCGSETSLKVNKLSEGHEYKFRVAVNRQGTSAPLTSD 3747  
QY 254 ----- 253  
:|  
Db 3748 HAIIVAKNPFDEPDAPTDTVPDWDKOHVDLEWKPPANDGGAPIDAYIVEKDKFGDWEC 3807  
QY 254 -----RKQAP----- 258  
:|  
Db 3808 ARVDGKTTKATADNLTPGETYQFRVAVNKAAGKPGSDPTGNVAVKPRRMAPKLNLAGLL 3867  
QY 259 ----- 258  
:|  
Db 3868 DLRIKAGTPIKLDIAFEGEPAPVAKWKANDATIDTGARADVNTTPTSSAIHIFSARVGD 3927  
QY 259 ----- 258  
:|  
Db 3928 GVYKIIIVENHKGDTAQCNVTVLDPGTPGGLKIDIEIHKEGCTLNWKPPDTONGGTDVLH 3987  
QY 259 -----VK 260  
:|  
Db 3988 YIVEKMDTSGTQWVGTPDCTAKVNKLVPGEYAFRVKANVLOGESKPLEABEPIIAK 4047  
QY 261 EQF-----EKKAK-----TQSTTTPY 277  
:|  
Db 4048 NOPDVPDPVKPEVTDWDKDRIDIKWNPANNGAPVTGYIVEKKEGSAIWEAGTTP- 4106  
QY 278 FSYTTSALNLVNTNVTY-----SITTAARV--- 303  
:|  
Db 4107 --GTFPSADNLKGEVEYFRVAVNNAAGSPDSDPTDPOITKARYLKPKILTASRKIK 4164  
QY 304 -----STSTIAY----- 310  
:|  
Db 4165 AGFTHNLEVDFIGAPDTATWVGSGAALAPPELLVDAKSSTISIFPPSAKRAADSGNYKL 4224  
QY 311 ----- 310  
:|  
Db 4225 KVNKLGEDEAIFEVIVQDRPSAPGPLEVSDTKDSCVLNWKPKDGGABEISNYVVEK 4284  
QY 311 ----- 310  
:|  
Db 4285 RDTKNTWVPVSFAVTGTSITVPKLTGEGHEYFRVMAENTFGRSDSLNTDPEVLAKDPFG 4344  
QY 311 ----- 310  
:|  
Db 4345 TFGKPRPEIVDTDNDHIDIKWDPDRNGSGSPVDHYDIERDKATGRWIKVNTSPVQGT 4404  
QY 311 ----- 310  
:|  
Db 4405 FSDTRVQKHTYFRVAVNKAAGCPQSDSSAAATAKPMHEAPKFDLDLKGKFRVKAGE 4464  
QY 311 -----RPDSSFWKS----- 319  
:|  
Db 4465 PLVITIPPTASPQPDISWTKGKPLAGVETDTSQTKLVIPTSTRSDGPVKIKAVNPYG 4524  
QY 320 ----- 319  
:|  
Db 4525 EABANIKITVIDKGPAPENITYPAVSRHTCTLNWDAPKDDGABAGYKIEYQVGSQIW 4584  
QY 320 ----- 319  
:|  
Db 4585 DKVPLGISGTAYTVRLEHGQOYFRIRAEANAVGLSDYCOQVPVVKDPDPPGAPSTPE 4644  
QY 320 -----IMATQ----- 324  
:|  
Db 4645 ITGYDNTQVSLAWNPPRDDGGSPLILGYVVERFEKRGGDWAPVMPVKGTECIVPGLHE 4704

QY 325 ----- 324  
:|  
Db 4705 NETYQFRVAVNAAGHGEPSEBVTCPYVEKEGAPDAPRVGKITKNSAELTWNRLR 4764  
QY 325 -----LRDLA----- 329  
:|  
Db 4765 DGAPIDGYIVEKKLGDNDWTRCNDKPEVRDTAFEVKNLGEKEBYFRVIAVNSAGEBP 4824  
QY 330 ----- 329  
:|  
Db 4825 SKPSDLVLEOPGRPIFDINNLDIIVRAGETIOIRIPYAGGNPKPIIDLFGNSPIFE 4884  
QY 330 ----- 329  
:|  
Db 4885 NERTVVDVNPGBEIVITTTGSKRSDAGPYKISATNKYKDTCKLNVFVLDAPGKPTGP 4944  
QY 330 -----TW-----VYTTLYR----- 339  
:|  
Db 4945 TDIQADAMTLSWRPPKONGGDAITNYVVEKRTPGGDWVTVGHPVGTTLVRNLDANTPYE 5004  
QY 340 -----QNPFCEP----- 346  
:|  
Db 5005 FRVRAENQYGVGEPLETDDDAIVAKNPFDTFGAPGQPEAVETSEAITLOWTRPTSDGAP 5064  
QY 347 -----SRNRTA----- 352  
:|  
Db 5065 IQGYVIEKREVGSTWTKAAGNILDTHKRVGTGLTPKTYEPRVAAVNAAGQGEYSVNSV 5124  
QY 353 ----- 352  
:|  
Db 5125 PITADNAPTRPKINMGMLTRDILAYAGERAKILVPPFAASPAPKVTFSGENKISPTDPRV 5184  
QY 353 ----- 352  
:|  
Db 5185 KVEYSDFLATLTIKSELTDGGLYFVELENSQSDSASIRLKVVDKPPASQHIRVEDIAP 5244  
QY 353 -----VSEFMKNTHVL----- 363  
:|  
Db 5245 DCCTLVMPSPSSDGGSPITNYIVEKLDLRHSDGKEKVSFVRNLNLYTVGGLIKONRYRF 5304  
QY 364 -IRNETPYTI----- 372  
:|  
Db 5305 RVRAETQYGVSEPCELADVVAKYQFEPVQPEAPTVRDKDSTWAELEWDPDPDGGSKII 5364  
QY 373 ----- 372  
:|  
Db 5365 GYQVQVTRDTSSGRWINAKMDLSEQCHARVTGLRQNGEFEFRIIAKNAAGFSKPSPPSERC 5424  
QY 373 -----YG----- 374  
:|  
Db 5425 QLKSRFPGPPIHVCASKSIGRNHCTIITWMAPLEDGGSKITGYNVVEIREYGSTLTWTVASD 5484  
QY 375 ----- 374  
:|  
Db 5485 YNVREPEFTVDKLREFNDEYFRVAVNAAGKGPISLPSGPIKIQSSGSGRPOIIVVKPBDT 5544  
QY 375 ----- 374  
:|  
Db 5545 AQPYNRRAVFTCEAVGRPEPTARWLRNGRELPESSRYRFEASDGVYKFTIKEVWDIDAGE 5604  
QY 375 -TLDMSLSLYNETMFV----- 389  
:|  
Db 5605 YTVVESNPNYSGDTATANLVQAPPVIEKDVNPVITLPSGDLVRLKLYFGSTAPFRHSLVLN 5664  
QY 390 -----ENKTASDSNKT----- 401  
:|  
Db 5665 REEIDMDHPTIRIVEFDDHILITIPALSVRAGRYEYTVSNDSGEATTGFWLNVLTGLPEA 5724  
QY 402 ----- 401  
:|  
Db 5725 PQGPLHISNIGBSTATLSWRPPVTDGGSKITSYVVEKRDLSKDEWVTVTSNKMNYIVT 5784



QY 402 -----PTSP-----SMGFO 410  
DB 5785 GLFENHEVEFRVAQNEGIGAPLVSEHPHIIARLPDPPTSPLEINLEIVQVGDDYVTLQW 5844  
QY 411 RTFIDPLWD-----  
DB 5845 R-----PLSDGGRLRGYIVKOEHEHDEWFRCNQPNPNVNPVNLIDGRKYRVFAV 5900  
QY 420 -----YLDLLF-----  
DB 5901 NDAGLSDLAELDQTLFQASGSGPKIVSPLSLENEEGRVTFECBISGSPRBYRWF 5960  
QY 427 -----  
DB 5961 GCKELVDTSKYTLINKGDKVLIINDLTSDDADEYTCRATNSSGTRSTRANLRIKTKPRV 6020  
QY 427 -----  
DB 6021 FIPPKYHGYEAQGETIELKIPYKAYPQGEARWTKDGEIENNSKFSITDDDKFATLRI 6080  
QY 427 -----LDEI-----  
DB 6081 SNASREDYGYRVVVENSGSDGTVNTVADVPEPRPFIENILDEAVILSWKPPALD 6140  
QY 431 -----  
DB 6141 GGLSVNTYTIKREAMGGSWSPCAKSRYYTYTIEGLRAGKQVEFRIIAENKHGQSKPCP 6200  
QY 431 -----  
DB 6201 TAPVLIPODERKRRRGYVDDEGKIVRGKTGVSSNYDNYVDFWQYTPQVPEIKHDVL 6260  
QY 431 -----RNFS-----  
DB 6261 DHDYIHEELGTGAPGVHVRTERATGNFNAKVFVTPHESDKETVKRIQTVSLRHPTL 6320  
QY 441 VNL-----  
DB 6321 VNLHDAFEDDNEVMVYIFBMSGGELFVKVADEHNKQSEDAVEYMRQVCKGLCHMHENY 6380  
QY 444 -----  
DB 6381 VHLDLKPENIMFTTKRSNELKLIIDGLTAHLDPKQSVKVTGTABFAPEVAEGKPVGY 6440  
QY 444 -----  
DB 6441 TDMNSGVLYILLSGLSPFGGNDDETLRNKVKSCDWNDDSAFSGISEDGKDFIRKLLL 6500  
QY 444 -----TTP-----  
DB 6501 ADPNTRMTIHOALEHPWLTPGNAPGRDSQIPSSRYTKIRDSIKTKYDAWPEPLPLGRIS 6560  
QY 447 -----EHR-----  
DB 6561 NYSSLRKHROPYISIRDAFWRSEAOFRIVKPYGTVEGOSANFYCRVIASSPPVTV 6620  
QY 450 -----RAVN-----  
DB 6621 HKDBELKQSVKMYKRYNGDYGLTINRVKGDGKEVTVRAKNSYGTKEBIVFLNVRHS 6680  
QY 454 -----  
DB 6681 EPLKPELEPMKAPSPRVEBEFKERRSAPFTFHLNRNLIQKHOCKLTCSLQGNPNT 6740  
QY 454 -----LST-----  
DB 6741 IEMKOGHPVDEDRVQVSRGVCSLEIFNARVDAGTYTATNDLGVDSVSECVLTQV 6800  
QY 454 -----  
DB 6801 KGGEPIPRVSSFRPRRAYDTLSTGTDVERSHAYDMMRRSLIRVSPDVRSAADDLTKI 6860  
QY 457 -----

DB 6861 TNELPSFTAQLSDSETVGGSAEFAAVSGQPEPLIEMHNGERISESDSRFRASVACK 6920  
QY 457 -----SNS-----  
DB 6921 ATLRISSDAKSDGOYLCEASASAGOEQTRATLTLYKGDOPLNGHAGQAVESELRVTKHL 6980  
QY 460 -----LM-----  
DB 6981 GGEIVNNGESVTFEARVQGTPEEVLMMRNGQELTNGDKTISQDGETLSFTINSADASA 7040  
QY 462 -----  
DB 7041 GHYQLEVRSGKTNLVSASLVVVGEXADPPVTRLPSSVASPLGSGTAFTIEPENVEGLTV 7100  
QY 462 -WW 463  
DB 7101 QWF 7103

## RESULT 11

T32650

hypochemical protein F39C12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T32650

R:Chisoso, S.; Sansone, J.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid F39C12.

A:Reference number: 221206

A:Accession: T32650

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-5105 &lt;CHI&gt;

A:Cross-references: UNIPROT:O61201; EMBL:AF039043; PDB: AAB94194.1; GSPDB:GN00028; CESP

A:Experimental source: strain Bristol N2; clone F39C12

C:Genetics:

A:Gene: CESP:F39C12.1

A:Map position: X

A:Introns: 19/2; 47/3; 103/2; 142/2; 174/2; 324/2; 362/3; 436/3; 494/1; 523/1; 923/3; 98

88/3; 4971/3; 5020/3

## Query Match

Best Local Similarity 19.3%; Score 472; DB 2; Length 5105;

Matches 213; Conservative 94; Mismatches 141; Indels 4284; Gaps 66;

QY 1 MGRKEMM-----VRDVPKM-- 14

DB 319 MLRREKMPYEDAEVPLLGYYGYEAANKNDPOVPIPTLWTPYAEVDMKFLAKIQDVPEIDY 378

QY 15 -----FVLISISFLIVSPINC----- 30

DB 379 ESYLYKAVERSAERDPKIHKPKILYQAKGVESDFV-----PFCVNGIMCFDRNTFTYK 433

QY 31 ----- 30

DB 434 IVRSPKHVGSVPYAGSDSVFFYRGLIQEIMAPIQTLTMNCRLEAGNAYQVRSERSLK 493

QY 31 -----KVMKALYN----- 39

DB 494 ELAPKITEERSVWAKLHNSVIIQYGGKEFGNIGVNDLEWDLVPLPDGRLPEVEIPPY 553

QY 40 -----RP----- 41

DB 554 EYMKDNVATEITOMVAPPFFLFLPDPRDLLSDSESETHNAPPNGNCCDDSDPEPSG 613

QY 42 -----WR-- 43

DB 614 LPSELFSNMRTLCKCHAKIRYNNKQSDRKVMYASMGNSPAIGSSKIRDKFFKLGWKS 673

QY 44 ----- 43

DB 674 TKQILGHLRKIDTAFKKKSIDLDLPKQFQKRPBQKRVLPPLKLTIDTSVFMKPEYIK 733

QY 44 -----GLVLSK----- 49  
 Db 734 EREAKEKEMREKAAOPTQAPIGLSMPKSKSRKEPKQTPVTSNGNEEDPEHARWIRAAIQ 793  
 QY 50 ----- 49  
 Db 794 LADNNSDDDEIVSPQATPTPPAPTPTPHVPOALSVRNEKDDPEYETPEDAYFGIKKR 853  
 QY 50 -----IGKVKLDOLK----- 59  
 Db 854 MYVYQATOPQLGLRRYKPDRLRESTFIRLEHLAMYGRRRHERDILYGRFEENDKROSS 913  
 QY 60 -----LEILROL-----ETT 69  
 Db 914 VKERTLDLKFSLNHNKVCWQTFYDKNFFSNKCSYNQILRKAIGMYNRLCGFDSHET 973  
 QY 70 I----- 70  
 Db 974 FKSVRHERMLCGVPRNVMDKEAIEBARYLKDISIENSTIKLYNHFTFVQAOLPYKL 1033  
 QY 71 -----STKN----- 75  
 Db 1034 DHTIFMENEGDTPFSQABCEPIEEKQTAKYDENDKFMPLGLRLRPFCEFDIPPFMHBI 1093  
 QY 76 ---VSKQPVKN----- 83  
 Db 1094 ENRRVPKPKCNPLGLEDHEHLSAQPSKPVLRSSNSCPDLGLAIESTENSRVCFNEFFHE 1153  
 QY 84 -----LTWNT 88  
 Db 1154 IYRRIIAYRDMCHQREHGNVPHYGSEYLEPLLTAGVLOQFNTHSVNGDYKVDDLTNFT 1213  
 QY 89 EF-----PO-----YYILAGPIQNS 104  
 Db 1214 KLEWDSPLSDTDRETGRKFRVRVPFKEPQPPQLEIGIDPLFYITIEVNRAAKNLH 1273  
 QY 105 IYLMFD-----F 112  
 Db 1274 VSYIFNDPILRTWPGILDEGRGMFTLCPKPPRPVPAQLSEIDI PKSPCLPTKAVEF 1333  
 QY 113 YS-----TQLRKPQKVVYQYNHTA----- 132  
 Db 1334 YSOFTEVLLIGRDSKKNRIRPPVYNEHNINVPGLIEILKPAFYGFYKPYDRESYNEEQ 1393  
 QY 133 -----KTIIFRPPPCGRVPSMTCLSEMLNVSK----- 159  
 Db 1394 SLEKNRRELQRAIFEKRIITFRQVIBETEQLAIEENLNVEEDDLTTSESEGEDEV 1453  
 QY 160 RNDT-----GROG----- 167  
 Db 1454 KNDSDPNDISDFWKIFKVTEDEAGVNGDRDSDDELVRKIGIFHAGDDSYNCFNPHTRE 1513  
 QY 168 ----- 167  
 Db 1514 NIRKISVBEVEDEPKLGSVLTNLFEQDLVKQRLGDCPIGSPRKSIIKIDYNRGMV 1573  
 QY 168 ----- 167  
 Db 1574 PLTELTKQTKVTWGEVTEHSVPPRPRIHRNRHRVEVDKPFVLDYSLVQLTRRVKNSDNF 1633  
 QY 168 ----- 167  
 Db 1634 QKTKPARKNSDPTAVHHIGIPRRASLDQFYRKDLPATCSACEKPEWPTKEICIVCHPDTT 1693  
 QY 168 -----CGN----- 170  
 Db 1694 IFRCTNVLNHEPCLICGNIIYKPCSNEYHNQVVTPIISIFDQPKLPVDEEFKKLA 1753  
 QY 171 ----- 170  
 Db 1754 VKNETVIYENTPAVMNVKAYRAGESAVNTFFAETDDRGFKTLQGDLAGFPHREFNVTW 1813

QY 171 -----FTTF----- 174  
 Db 1814 RMEDEKSERVFVSEBFDVAFDRDADIDMMRDTMTNISODGEMDMWTTVDQRTGMRIRRWI 1873  
 QY 175 ----- 174  
 Db 1874 SYAALATRVQKQNEKVHYRKKAAEREARILSGEISVPSLNLNELKKAKOKEVAATIATE 1933  
 QY 175 ----- 174  
 Db 1934 NDDAKKEERRKAYKWKLPPIPEFIEDPKKKPNFAPPKSEDMLESSGSEDELPPIKSR 1993  
 QY 175 ----- 174  
 Db 1994 ILSKEMKEFIDTIRVHMEQKPSFEGIAVYGLFEKRPILKNDIVLVRKKNVRRKPPP 2053  
 QY 175 -----NPMFFNVPRWNT----- 186  
 Db 2054 FRIQRIPLDESMLNATFTVSPFLVPVQVRNPRDPMKWTVPFYNQORLALTAQETKERALR 2113  
 QY 187 -----KUYV-----GPT 193  
 Db 2114 PPTPPPTLKKLYTTMKDVEKAFYHRKRVKSOITLLKAKTARQLRHFHKLYGHSKEPSHPT 2173  
 QY 194 KV----- 195  
 Db 2174 KAWTHVRQIGLKKSKVQPRKLEAHYLPGLAKOKNKYIPTYFPNPKLEKVKLSQRKPSR 2233  
 QY 196 ----- 195  
 Db 2234 ISSASDSEYCPIVKFPKQKRVVELKEDHPNNSSELGIERKKELKREDDPPAVLMSSTR 2293  
 QY 196 -----NVDSQT----- 201  
 Db 2294 STLKRPETGSFVREISVGRKPALFQKADVQKLTETVSLKRIDSQTEPKTESIKORRHIFE 2353  
 QY 202 -----IYFLGLTAL----- 210  
 Db 2354 HLEDKSIINFSQIGIPALISVVEPSSCOLPRKSRQKEHKITNNSNSNISIDSDGMMDRSS 2413  
 QY 211 -----LLRYAQKNC----- 219  
 Db 2414 KQPSVAVRSKSTMIPIHSRKLKQMPACRSPSVDKSGSYRKALKIARKKCRSIOKEKSKV 2473  
 QY 220 ----- 219  
 Db 2474 TKPDDSKPEVIPSQKPKSQLETSVALPSKSAKVSTKSEKOLSPKSTELQLERKFLKN 2533  
 QY 220 -----THSFYL----- 225  
 Db 2534 KVHNRSSRAPGIHHQRPPLHPTDHDTRDKDSRAASQFRFQHMRSKRFTQKVLHK 2593  
 QY 226 -----VNAMSRNLFRVP-----KYIN 241  
 Db 2594 EPRRSVLNITIRKNDLSLKGKHNHMKYEQPIDAKISDMFLPKMPKPLRRSKKIRSKKYYK 2653  
 QY 242 GTKLKNT-----MRK----- 251  
 Db 2654 NVKAKNTTKLOPMKPKQMLKRLADYNNIKRVPLSDFKIRPKQOEDMETKEKSIIVVK 2713  
 QY 252 -----LKRQ-----APVKEQF-- 263  
 Db 2714 PSNNLKSCLKRGQGHRSKMFNSQSPDGKTSAGSNKTLKLVKTSAGHLAGTSPPKTSFTV 2773  
 QY 264 -----EKKAKK----- 269  
 Db 2774 PSLELDDSSPLKNVLTWLTNRKNSKGEKVKENTAPVQOAPVPVSPPRETPEPTSPPV 2833  
 QY 270 -----TOSTTTP----- 276  
 Db 2834 RLLRPLSABSEKVPVEKVDTPSPLTPFAASVPAPPPKMPADQOQWRAAMWARRYMCAKSP 2893  
 QY 277 -----YFSYTT----- 282

Db 2894 VFVKPSALSATESLKEASNTQESASTTRVQPALVSHSTPEPVQLTRMAAKKSEQKEY 2953  
QY 283 -----SALNV----- 288  
Db 2954 KTPVAKPKATAPPITPKAASIRATQKLHESELTKSEASRNVSSSGSKSQKQADVQSRP 3013  
QY 289 -----TTNVY----- 294  
Db 3014 KLSQVQKRSVPKVVKPLDPKPGQKQESKNEKVPSTNVTLDDQASSSNQOGPSAK 3073  
QY 295 -----ARRVST----- 307  
Db 3074 KPSASITFALFGRKKEKMTSSTAESSIPISPENQEAETSNRSPILPRIKPKETGQ 3133  
QY 308 ----- 307  
Db 3134 LIKESPTSDVPKPCVSGAIVKFDGKPISNSPSYKGMARRPKKADSAADSGDLSDE 3193  
QY 308 -----IAYRP----- 312  
Db 3194 GSTSHEGCLKDEENSKKQKAGKTVRFGPISYEPPAPAPQVEIPKVEPTPEISFESLAQ 3253  
QY 313 ----- 312  
Db 3254 QMAQKTLRKEASKILEHPVTEQTSPTABVPSFTKSGSKLSPFGLSPFGKVTPKQNTA 3313  
QY 313 ----- 312  
Db 3314 QTQSLTVLESTSSPELLTPSNPETSNTASTSSSKQPKSLSEBKITATQMIQTRGRPIRKL 3373  
QY 313 -----DSS----- 315  
Db 3374 SHRQLKIGDPSPLQLSGDSSLRPAQPSPEKPPQLEVPGVQVQYNNPPSQRYVRLGRE 3433  
QY 316 ----- 315  
Db 3434 WKPPPKAKTPLSIRVQSGVSTSPAIDLSAKEVLLAPIEKTQKQKRPKPMETTLVLVD 3493  
QY 316 ----- 315  
Db 3494 GTEHFKLTAENQLLSRQFKSKRKNKKTWTSVIGCDENDSTKQMHKLLBEFRD 3553  
QY 316 FMKSMATOL----- 325  
Db 3554 LKKSLLITRQDKMITARAVNPRFTDSSSTKHELLRNKPKFTTRKCLKELANNRQKI 3613  
QY 326 -----RDL-----ATWY- 333  
Db 3614 RMSGIHQPKRDVAISAIROVKVRFPQSTGKGMNPIISGFKPKSILKPKPPATWLF 3673  
QY 334 ----- 333  
Db 3674 IDSEGNLYMKLGPQDWQKQNDKCNPMGANDPRKKSNSYLKRWAKQMAKVGKQNTK 3733  
QY 334 ----- 333  
Db 3734 LFETNQLTRRLGAPKDSHPKDLVDPKADAPPPSRSLSWTTVSSNASLSLLDALNTV 3793  
QY 334 ----- 333  
Db 3794 GSDEPSTSGIVTDIAPPIVKEQPRKAKESRPTQOTVSKYPTVYLGMTVARSPSSSTSL 3853  
QY 334 -----TTLRYONPCEP----- 346  
Db 3854 VNLDERAVSSLAQDSEIVEYVQBEPPVETTKRRFRDPFVRPISALFKAPTSSSPEVGL 3913  
QY 347 -----SRNR-----TAVS----- 354  
Db 3914 EADSRSPRTERTLGLFRSSPTMPESTRHVSSPTERRANRGPVTSVQSSAEDILA 3973  
QY 355 ----- 354

RESULT 12  
I38346

Db 3974 LMOAQPDALENMBEQVSVKPPPIQLQAPVLNKGSMWRPKLKNKLSLYTQFKELKEESKUR 4033  
QY 355 -----EFMK-----NTHVLIRNETPYTYTLD 377  
Db 4034 KLVEYLDQNYKNKSYSTLQTLITFNLDQYMRTNLYPRPKLNTDLL-----GKYD 4083  
QY 378 MSSL----- 381  
Db 4084 PNSLRVSSSSRGLQRSTRLVKIRRCORQLEKWKELIATSTNQQTNGELVKRAKAM 4143  
QY 382 ----- 381  
Db 4144 QKASNRKAKKVLVLYKFKRTCKFRVQVTTTKSARKILADONQKKQLIMHEEKRTPTIL 4203  
QY 382 ----- 381  
Db 4204 TIGDEDDENEIVDLLSGNEDSANQNLADYDGDNDDDAIDTELDWVANTSQVIRKQRR 4263  
QY 382 -----YVN----- 384  
Db 4264 RLLRHYFNFAIRMARKVSRRSVGQKPLVFYSFISQMSRMRNPMRIRHNDQTRARQNK 4323  
QY 385 -----ETMFVENKTASDSNKTPTSPS-----NGFQ 410  
Db 4324 LAVSNSTTLKETRLRSSKKQNRKSKSQROIPICIIRNSLKYGGTAEKKVIVEOSKFNQ 4383  
QY 411 R-----TF-----IDPLM----- 418  
Db 4384 KLATKLOTFFECLEIDVGTPIKLVFSSKNTILPKQWFDVKTCLIDPVYVTSPCNTD 4443  
QY 419 -----DYLSSL-----FLDE----- 429  
Db 4444 SDFSDNTDDYKSDLKQFPLRESLESTYHVKGGLIYLARLVQERVADDVERNIIPRG 4503  
QY 430 -----IRNFSLSRP----- 438  
Db 4504 PGMETCNLVVYMTPIRRLMLNSFGLRMPDRLLDVMLSKHQNEKXPELFDRLYMLCYGRG 4563  
QY 439 -----TY----- 440  
Db 4564 DLCKYIDKISLQGPNNKYEKMDITYQAHIIHKLYKQMYIENGREAKLAVQEFMDIG 4623  
QY 441 ----- 440  
Db 4624 LDIDHEPFSKQVKMGQYDRMLYGNPKPNKDNWVEACDQFLATYNTKEDVLVSCQASILL 4683  
QY 441 -----VNLTPPEH----- 448  
Db 4684 MRMIPLGGYSQDVVDHLITWNLPMRDKLFIINLTPDNLVENIVKIFHTVVRGLDRLPFV 4743  
QY 449 -----RRA----- 451  
Db 4744 GLGYHADPLRVLFEEAVESGDPQLIAHIVRGRCLDRSRPEMPTDCGSDHIQKDFANI 4803  
QY 452 ----- 451  
Db 4804 MTKVITNGNERRPYGAPTMDEMPFRYVRNPDGSDVAETTEELANIMECEVTRYMAILE 4863  
QY 452 -----VNLSTNSL----- 460  
Db 4864 TKGAAVTKGDLHNPKNTYPHPKPYFDVLCTGCMNVSKGRNLGRGLRTLCVSTMPLEIQ 4923  
QY 461 -----W----- 461  
Db 4924 DXDKKRWPNKMTSPRLTPENYIETLPTENKMLMAKIDBEKTDITVQVLPSPFGCFCQ 4983  
QY 462 ----- 463  
Db 4984 RPSLPRCCVCMCPYLNHDHMDFALKDYNESFQPSICNICNHGGHVNHIATWF 5035  
-----W 463  
----- 461

elastic titin - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 09-Jul-2004  
C:Accession: I38346  
R:Label: S.; Kolmerer, B.  
Science 270, 293-296, 1995  
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.  
A:Reference number: A57430; MUID:96026330; PMID:7569978  
A:Accession: I38346  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-7962 <RES>  
A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017426  
C:Genetics:  
A:Gene: GDB:TTN  
A:Cross-references: GDB:127867; OMIM:188840  
A:Map position: 2q31-2q31

Query Match 19.2%; Score 471; DB 2; Length 7962;  
Best Local Similarity 2.7%; Pred. No. 3.7;  
Matches 211; Conservative 91; Mismatches 154; Indels 7235; Gaps 68;

QY	2	GR-----	3
Db	158	GREIYSDKCSIRSKYISSLEILRTQVDCGEYTCASNEYSGVSCATLATVTEAYPPT	217
QY	4	-----	3
Db	218	FLSRPKSLTTFVGKAAFICTVGTGPIETIQKDGAAALSPSPNWRISAENKHILELSN	277
QY	4	-----KEM-----	6
Db	278	LTQDRGVSKASNKFGADICQAEIILDKPHFIKELEPVQSAINKVHLBCQVDEBK	337
QY	7	-----M	7
Db	338	VTVMSKDGKLPFGKYKICPEDKIATLEIPLAKDKSGTVTCASNEAGSSSCSATVT	397
QY	8	VRDVPKMFVLSISFLV-----	25
Db	398	VREPPFVKVDPDSYLMPLFESARLHCKLKGSPVQVTFWFKNNKELSESNTVMYFVNSE	457
QY	26	-----SPT-----	28
Db	458	AILLDITDKVEDSGSYSCAVNDVSGSDSCSTEIVIKEPPSFIKLEPADIVRGTNALLQC	517
QY	29	-----NC	30
Db	518	EVSGTGPPFISWFKDKQIRSSKYRLFQSKSLVCLEIFSNSADVGVECVVANEVGC	577
QY	31	KVMSKALYNRP-----	41
Db	578	GCMATHLLKEPPTFVKVDDLLALGGQVTLQAAVRGSEPISTVMKRGQEVIREDEGKIM	637
QY	42	-----	41
Db	638	SFSGNGAVALLIPDVQISFGKYTCIAENEAGSQTSVGLIYKEPAKIIERAEILQVTAGD	697
QY	42	-----W-----RGLVLSKIYKYL-----OLK-----	59
Db	698	PATLEYTVAGTFELKPKRYKDGRLVASK--KYRISFKNVNAQLKPYSAELHDSGQVTP	755
QY	60	-----	59
Db	756	ISNEVGSSSCETTFTVLDRDIAFFTFLRNVDVNVGTCLRDLCKIAGSLPMRVSWFKDG	815
QY	60	-----LEILR-----	64
Db	816	KEIAASDRYIAFVEGTASLEIIRVDMNDAGNFTCRATNSVGSKDGGALLIVQEPFSFT	875
QY	65	-----	64
Db	876	KPGSKDVLPGSACLKSTFGQSTPLTIRWFKGNKELVSGGSCYITKEALESLELYLVKT	935

QY	65	-----	66
Db	936	SDSGTYTCKVSNVAGVGCSSANLFVKEPATFVEKLEPSQLLKGDATQLACKVTGTPTPIK	995
QY	67	-----	66
Db	996	ITWFANDREIKESSKHRMSFVESTAVLRLTDVGIEDSGEYCEAQNAGSDHCSSIVIVK	1055
QY	67	-----	66
Db	1056	ESPIYTFKEFKPIEVLKEYDVNMLAEVAGTPPEITWFKDNTILRSGRKYKTFIQDHLVSL	1115
QY	67	-----	66
Db	1116	QILKFAADAGEYQCRVTNEVGSSICSAVTLRPPSPFIKIESTSSLRGGTAFAQATLK	1175
QY	67	-----	66
Db	1176	GSLPITVTLKDSDEITDDNIRMTFENNVAISLYLGGIEVHDKYVQCAKNDAGIQRCS	1235
QY	67	-----	77
Db	1236	ALLSVKEPATITEAVSIDVTQGDPAQLVKFGTKEITAKWPKDGOELTLGSKYKISVT	1295
QY	78	-----KQPVKNLTM-----	86
Db	1296	DTVSILKIIISTEKDKSGEYTFEVDVGRSSCKARINVLDLIIPPSFTKKLKWDISIGS	1355
QY	87	-----NTEP-----	90
Db	1356	FIDLECIAGSHPISTQWFKDDQEIASEKYPFSFHDNTAFLEISQLEGTDSGTVTCSAT	1415
QY	91	-----POYYI-----	95
Db	1416	NKAGHNCQSHLTVKEPPYFVEKPSQSDVNPNTVOLKALVGTAPEWTIKWPKDNKELHS	1475
QY	96	-----	95
Db	1476	GAARSVMKDDTSTSLBELFAAKATDSGTIYICQLSNDVGTATSKATLTVKPEPQFIKKPSFV	1535
QY	96	-----	95
Db	1536	LVLRNQSTTFEQITGTPKIRVSWLDGNEITAQKHGISFIDGLATFQISGARVENS	1595
QY	96	-----	95
Db	1596	TVCEARNDAAGTASCIELKVKKEPPTFIRELKPVEVVKYSDVELECEVTGTPPEVTWLK	1655
QY	96	-----	95
Db	1656	NNREIRSSKYYTLTDVRSVFNHLITKCDPSDTGEYQCVISNEGGSCSSTRVALKBPSPF	1715
QY	96	-----LAG--PIQNYISITVL-----	108
Db	1716	IKKIENNTTLKSSATFQSTVAGSPPI--SITWLKDDQILDEDDNVVISFVDSVATLQI	1772
QY	109	-----	108
Db	1773	RSVDNGHSGRYTCQAKNESGVERCVAFLLVOEPAQIVEKAKSDVDTEKDPMTLECVVAGT	1832
QY	109	-----	108
Db	1833	PELVKWLKDGQIVPSRYFSMSPENNVAISFQSVMKQDSGOYTFKVENDFGSSSCDAY	1892
QY	109	-----	108
Db	1893	LRVLQDNIPTFTKLTMDKVLGSSIHMECKVSGSLPISQWFKDGKEISTSAKYRLVC	1952
QY	109	-----	108
Db	1953	HERSVLSLVNLELDTANTYCKVSNVAGDDACSGILTVKEPSPFLVKPQOQAIIPDSTV	2012

QY 109 -----WF-----DFYST----- 115  
Db 2013 EFKAILKGTBPKIKWFKODVELVSGPKCFGLGEGSTSLNLYSDVASKTGQYCHVTND 2072  
QY 116 ----- 115  
Db 2073 VGSDCSTMLLVTEPPKFKLEASKIVKAGDSSRLECKIAGSPEIRVVWFRNEHELPA 2132  
QY 116 ----- 115  
Db 2133 DKYRMTFIDSVAVIQMNNLSTEDSGDFICEAQNAGSTSCSTKVIVKEPPVFFSPPIVE 2192  
QY 116 -----OLRPAKYYSQYN-HTA----- 132  
Db 2193 TLKNAEVSLECELSGTPPFVWVVKDKRQLRSKKYKTIASKNPFHTSIHLNVDTSDIGEY 2252  
QY 133 ----- 132  
Db 2253 HCKAQNEVSGDTCVCTVKLKEPPRFVSKLNSLTVVAGEPABEQASIEGAQPIFVQWLKEK 2312  
QY 133 ----- 132  
Db 2313 EEVIRESENIRITFVENVATLQAKAEPANAGKYICQIKNDGGMENMATLWLEPAVIV 2372  
QY 133 ----- 132  
Db 2373 EKAGPMTVTVGCTCTLECKVAGTPELSVEMYKDGKLLTSSQHKRFSFYNNKISSLRILSVE 2432  
QY 133 ----- 132  
Db 2433 QODAGTYTFQONNVGKSSCTAVDVSDRAVPPSFTRRLKNTGGVILGASCILECKVAGSS 2492  
QY 133 ----- 132  
Db 2493 PISVAMFHEKTI VSGAKYQTTFSNDVCTQLNSLSDSDMGNYTCVAANVAGSDECRV 2552  
QY 133 -----KTIITFR-----PPP----- 141  
Db 2553 TVOBPPSFVKEPPELVLPGKNVTFTSVIRGTPPFKVMWFRGARELVKGDRCNIYFEDTV 2612  
QY 142 ----- 141  
Db 2613 ABELFNIDISQSGEYTCVSNNAQASCTRLRVKEPAFLKRLSDHSVEPGKSIILES 2672  
QY 142 -----CGRVP-----SMTCLSEMLNYSKR-----NDTGEQ 166  
Db 2673 TYTGTLPISVTWKXGDNFTTSEKCNIVTTEKTCILEILNSTKRDAGQYSCIEENAGRD 2732  
QY 167 CGGNF-TTFNPMFP-----NVPWNTKLYVGP----- 193  
Db 2733 VCGALVSTLEPPYFVTELEPLEAAVGDVSLQCQVAGTPEITVSWYKGDTKLRPTPEYRT 2792  
QY 194 -----KUNVD-----SOTIY----- 203  
Db 2793 YFTNNVATLVFNKNVINDSGBYTCKAENSIGTASSKTVFRIQERQLPPSFARQLKDIEQT 2852  
QY 204 ----- 203  
Db 2853 VGLPVTLCRLNGAPIQVCHVRDGVLLRDHENLQTSFVDNVATLKILQTLDSHGQYSC 2912  
QY 204 ----- 203  
Db 2913 SASNPLGTASSARLTAREPKKSPFFDIKPVSIDVIAGESADFECHVTGAQPMRITWSKD 2972  
QY 204 ----- 203  
Db 2973 NKEIRPGNYTITCVGNTPHLRILKVGKDSGQYTCQATNDVGMKSAQLSVKEPPKFV 3032  
QY 204 -----FLGTALL----- 212  
Db 3033 KXLEASKVAKOGESIQLECKISGSPKIVSWFRNDSSELHESWKYNMSPINSVALLTINEA 3092  
QY 213 ----- 212

Db 3093 SAEDSGDYICEAHNGVGDASCSTALTAKAPPVFTQKSPVGCALKGSVDILOCEISGTPPF 3152  
QY 213 ----- 212  
Db 3153 EVWVKDKOVNRSSKPKITSKHFDTNLHILNLEASDVGEYHCKATNEVGSDTSCSVK 3212  
QY 213 -----RYAQR----- 217  
Db 3213 KEPPRFVKCLSDTSTLIGDAVELRAIVEGFQPISVWMLKDRGEVIRESENTRISFIDNIA 3272  
QY 218 ----- 217  
Db 3273 TLQGSPEASNGKYICQIKNDAGRCSAVLTLEPARIIEKEPEMTVTTCGNFPALECV 3332  
QY 218 ----- 218  
Db 3333 VTGTPELSAKWFKDGRELSDSKHHITPINKVASLKIPCAEMSDKGLYSPEVKNSVGKSN 3392  
QY 219 CTHSFYL-----VNAM----- 229  
Db 3393 CTVSVHVSDRVPPSFIRKLKDNVAILGASVLECRVSGSAPISVGMFQDGNBIVSGPKC 3452  
QY 230 -----SRN----- 232  
Db 3453 QSSFSENVCTLNLSLLEPSDTGIYTCVAANVAGSDECSAVLTVOBPSPFEQTPDSVEVLP 3512  
QY 233 -----LPRV----- 236  
Db 3513 GMSLTPTSVIRGTPPFVKWFKGSRLELVPGESCNI SLEDFTVELELEFEVQLESQDYSCL 3572  
QY 237 ----- 236  
Db 3573 VTNDAGSASCTTHLVFVKEPATFVKRLADFVETGSPIVLEATYTGTPPISVSWIKDEYLI 3632  
QY 237 ----- 236  
Db 3633 QSERCISITWETKSTILEILESTIEDYQAQYSLIENAGQDICEALVSLVPPFIEPLE 3692  
QY 237 -----PKY----- 239  
Db 3693 HVEAVIGEPAATLQCKVDGTPAIRISWYKEHTKLSAPAYKMQPKNNVASLVINKVDHSDV 3752  
QY 240 -----ING----- 242  
Db 3753 GEYSCKADNSVGAVASSAVLVIKARLPPFPFARKLDVHETLGPVAFECRINGSSEPLQV 3812  
QY 243 ----- 242  
Db 3813 SWYKQGVLLKDDANLOTSFVHNVATLQILOTDQSHIGQYNCASNPLGTASSAKLILSE 3872  
QY 243 -----TKLKXNT----- 248  
Db 3873 HEVPPFPDLKPVSDVLDLALGESGTFKCHVTGTAPIKITWAKDNREIRPGNYKMTLVENTA 3932  
QY 249 ----- 248  
Db 3933 TLTVLKVGKDAGQYTCVASNIAGKSDSCSAQLGVQEPPEPRFIKLEPSRIVKQDEFTRVEC 3992  
QY 249 ----- 248  
Db 3993 KIGGSPEIKVLWYKDETEIQESSKFRMSFVDSVAVLEMHNLSDVSDGDTCEAHNAAGSA 4052  
QY 249 --MEKLARKQAPVKEQFEK----- 266  
Db 4053 SSSTSLEKVEPPI---FRKKPHIETULKGADVHLECEBLOGTTPPHVSWYKDRLERSKK 4109  
QY 267 ----- 266  
Db 4110 YKIMSENFLTSIHLNVDAADIGBYQKATNDVGSDDTCVGSIALKAPPRFVKKLSDISTV 4169  
QY 267 -AKTQSTTT----- 275

Db 4170 VQKEVQLQTTIEGABPISVWFKDKGEIVRESDNWISYSENATLQFSRVEPANAGKYT 4229  
 QY 276 ----- 275  
 Db 4230 CQIKNDAGMQCBFATLSVLEPATIVEKPEISKVTTGTCTLECTVAGTPELSTKWFKDGK 4289  
 QY 276 ----- PVP 278  
 Db 4290 ELTSDNKYKISFFKNVSGLKIINAVPSDSGVYSFVQNPVKGDSCTASLQVSDRTVPPSP 4349  
 QY 279 S----- 279  
 Db 4350 TRKLKETNGLSGSSVWMECKYVGPPISSVWFHEGNEISSGRKYQTTLDNTCALTVNML 4409  
 QY 280 -----TSALNVT-----TNVTYS----- 295  
 Db 4410 EESDSGDTCTIATWAGSDECSAPLTVREPPSPVQKPDMDVLGTGNVTFTSIVKGTPTPP 4469  
 QY 296 -----ITTAARRVSTSTIAY- 310  
 Db 4470 SVSWFKGSSELVPGDRCNVSLSDSVALELFDVDTSGSGEYTCIVSNEAGKASCTHLYI 4529  
 QY 311 ----- 310  
 Db 4530 KAPAKFVKRLNDYSIEKGKPLILEGFTGTPPISVTVKXNGINVTSPQRNITTEKSPI 4589  
 QY 311 ----- 310  
 Db 4590 LEIPSTVEDAGQVNCYNIENASGKSCSAQILILEPPYFKQLEPVKVSVDGASLQCOL 4649  
 QY 311 -----RPDSSF--WKSIMAT----- 323  
 Db 4650 AGTPEIGVSWYKGTDLKRLPTTKYKMFHFRNNVATLVFNQVDINDSGEYICKAENSVEVSA 4709  
 QY 324 -----QLRDL-----ATW----- 331  
 Db 4710 STFLTVOEQKLPSPFSRQLRDVQETVGLPVVFDCAISGSEPISVSWYKDGKPLKSPNVQ 4769  
 QY 332 -----VYTTLR----- 337  
 Db 4770 TSPFLDNTATLNFKTDRSLAGVSTATNPISGASSARLIITEGNPPFPDIRLAPVDA 4829  
 QY 338 ----- 337  
 Db 4830 VVGESADFCHVTGTQPIKVSNAKDSREIRSGGKYQISYLENSAHLTVLKVDKGDGQYT 4889  
 QY 338 -----Y 338  
 Db 4890 CYAVNEVGKDSCTAQLNIKERLIPSPFKRLSETVEETEGNSFKLEGRVAGSQITVAVY 4949  
 QY 339 RON----- 341  
 Db 4950 KNNIBIOPTSNCEITFKNNTLVQVRKAGMNDAGLYTCKVNDAGSALCTSSIVIKBPKK 5009  
 QY 342 ----- 341  
 Db 5010 PPVFDHLPVTVSEGEVQLSCHVQSGSEPIRIQWLKAGRIKPSDRCSFSGTAVLE 5069  
 QY 342 -----PFCEPSNRNRTAV----- 353  
 Db 5070 LRDVAKADSGDYCKASNAGSDTTKSKVTIKDKPAVAPATKKAADVGRLLFFVSEPOSIR 5129  
 QY 354 ----- 353  
 Db 5130 VVEKTATFIKVGGDPIPNVKNVTKGWRQLNOGGRVFIHQKGEAKLEIRDTTKTDSGL 5189  
 QY 354 ----- 353  
 Db 5190 YRCVAFNEHGBIESNVNLQVDERKKQEKIEGDLFAMLKTKTPIILKKGAGEEBEIDIMELK 5249  
 QY 354 ----- 353  
 Db 5250 NVPKPEYKARMYGITDFRGLLOAFELLKQSOBEETHRLEIBEIERSEDERKEFEELVS 5309

QY 354 ----- 353  
 Db 5310 FTQORLSQTEPVTLLKDIENQTVLKONDVAFEIDIKINYPEIKLSWYGTGTEKLEPSDKFE 5369  
 QY 354 ----- 353  
 Db 5370 ISIDGRHRTLRYKNQCKDQGNRYLVCGPFIASAKLTVIEPAWHERHLQDVTLKESQTCMTM 5429  
 QY 354 -----SEPMKNTHL----- 363  
 Db 5430 TVQFSVPNVKSEWFRNGRILAKPQGRHKTEVEHKVHKLTADVRAEDQGYTCYKEDLETS 5489  
 QY 364 ----- 363  
 Db 5490 AELRIEABPIQTKRIQNIWVSEHQSATCEVEVSFDDAIVTWYKGPTELTESEKYNFRND 5549  
 QY 364 -----IRNETP-----YTIYGTI----- 376  
 Db 5550 GRCHWMTIHNVTDDDEGVYSVIARLEPRGEARSTAEYLTLTKEIKLEKPPDIPDSRVPI 5609  
 QY 377 ----- 376  
 Db 5610 PTMPIRAVPPEEIPPVVAPPVPLLLPTPEKKPPPKPIEVTKKAVKDKAKVVAKPKEMT 5669  
 QY 377 -----DMSL----- 381  
 Db 5670 PREBIVKPPPTTLIPAKAPEIIDVSSKABEVKIMTITRKEVQKEKEAVEYKQAVHK 5729  
 QY 382 ----- 381  
 Db 5730 EKRVIESPPEPDELEVBYTTPPEQPYEEDDEYEEIKBAKKEVHEEEDFEQG 5789  
 QY 382 -YY----- 383  
 Db 5790 EYTEREGYDEGEWEAEAYQEREVIQVQKEVVEESHERKVPKAPKPEKAPPPKVIKXP 5849  
 QY 384 -----N 384  
 Db 5850 VIEKIETSRMBEEKQVTKVPEVSKKIVPQKSPRTPVQEEVIEVKVPVHTKKNVISE 5909  
 QY 385 ETMFVENKT----- 393  
 Db 5910 EKMPFASHTEEVSVTVPQKEIVTEEKIHVAKSRVEPPPKVPPELPEKPAPEEVPAPV 5969  
 QY 394 ----- 393  
 Db 5970 IPKKVEPPAPKVPVPPKVPBEKKVPVPVPPKPEAPAPKVPVPPKVPPEEKI PVPVACK 6029  
 QY 394 ----- 393  
 Db 6030 KEAPPAKVPVQGVVTEESKIITVTORESPPPAVPEIPKPKVPBEKVPVPRKBEVEPPP 6089  
 QY 394 -----ASDSNKT----- 400  
 Db 6090 PKVPALPKVPPEEKVAVPVVPAKKAPPRAEVSCKTTVVEEKRFFVAEEKLSPAVQORVEV 6149  
 QY 401 ----- 400  
 Db 6150 TRHEVSAEEMSYSBEEEGVSVSVYREBEREEBAEYVEVMEBEPBEYVVEEKLHIIS 6209  
 QY 401 ----- 400  
 Db 6210 KRVEAPEAVTERQEKIVLKPKIPAKIEBPPPAKVPAPKPIVPEKKVPAPVPPKKEVP 6269  
 QY 401 ----- 400  
 Db 6270 PPKVPEEPKVPVPEKKVPPKVIKMBEPLPAKVTKEHMQITQEEKVLVAVTKKEAPPKARV 6329  
 QY 401 ----- 400  
 Db 6330 PPEPKAVPEEKVLKLPKREBPPKVPTEPRKRVVKEKVSIEAPKREPOPIKEVTIME 6389

```
QY 401 ----- 400
Db 6390 EKERAYTLTEEAASVORSEEEYSEYDYKBEFEYPTTEYDQYBEYEREYEEHEE 6449
QY 401 -----TPTSP----- 405
Db 6450 YITEPEKIPVKEPVEEPVPTKPAKPAKVLKKAIVEKVPVPIPKLKPDPKVPBP 6509
QY 406 ----- 405
Db 6510 KVFEKIHISITKREKEQVTEPAAKVPMKPKRVAAEKVPVPRKEVAPPVVRVPEVKELE 6569
QY 406 --SMGFQRTFDLWDYL----- 421
Db 6570 PEEVAFBESEVTHVEYLVESEESIHEEBSFITSEEVVPIPVKVPVPRKPVEEKKP 6629
QY 422 ----- 421
Db 6630 VPVPKKEAPKAPVPEVKKEEKVPVLI PKKEKPPKAPVPEVKVPPEKVPVVPK 6689
QY 422 ----- 421
Db 6690 VEAPPKVPPEVKVPVPAKVPKVPKVEAPPKVPVPKLIPEEKKPTVPKKEVAP 6749
QY 422 -----DSLLFLDEI----- 430
Db 6750 PPKVPKREVPVPVVALPQSEEVLPQSEEVLPVEEVLPEEVLPEEVLPEEVLPEE 6809
QY 431 ----- 430
Db 6810 EEIPEEVEVPPEEYVPEEVEEVEEVLPEVKVPVPAKVPVPEIKKVKTEKKVVPK 6869
QY 431 ----- 430
Db 6870 EEAPPKVPPEVKVEEKRIILPKKEEVLPEVTEPEEPISEEEIPEEPPSIEEVEE 6929
QY 431 ----- 430
Db 6930 APPRVPEVIKAPVPAEPTVPVKVEAPPKVKIPEKVPVVPVKKEAPPKVPPEVK 6989
QY 431 ----- 430
Db 6990 VPEKKVLVPKKEAVPPAKGRTVLEBKVSAPRQEVVVKERLELEVAAVEEIPPEE 7049
QY 431 ----- 430
Db 7050 EVEEYFESEGFHEVEEFTKLQHRVBEHREKVRHVLVEFEAEVEVEFEKPKAPPK 7109
QY 431 ----- 430
Db 7110 ISEKIIPPKDPTKVPRKEPPKAPVPEVKIIVVEEKVRVPEEPRVPPTKVPEVLPP 7169
QY 431 ----- 430
Db 7170 VPEKKVPVPAKKEAPPKVPPEAPKVEVPEKKVPVPPPKPEVPTKVPEVPAKAAV 7229
QY 431 ----- 430
Db 7230 KVPEAIPPKPESEPPPEVEEESPSAPPKKEVPPVVRVPEVKVEKVPKPAAPK 7289
QY 431 ----- 430
Db 7290 EVTPVKVPEAPKEVVEKKVPVPPPKPEVPTKVPEVKVAPVEKKVPEAIPPKPE 7349
QY 431 ----- 430
Db 7350 PEVFEPEBEVALEBPAAVEEPEAPPPQVTPPKNPVPEKAPAVVAKPELPPVKVP 7409
QY 431 ----- 430
Db 7410 EVPKEVPEKKVPLVVPKPEAPPKAPKVPPEVVEKVAVPKPEVPPKAPVPEVKP 7469
QY 431 ----- 430
```

```
Db 7470 VLEBKPAVPVERAESPPEVYVEBPEETAPBEEIAPBEEKVPVVAESEEPEVPPPAV 7529
QY 431 -----RNFSLSRS 437
Db 7530 PKKIPEKKVPVVIKKPEAPPPKEPEKVIKPKLKRPPPPPPAPPKEDVKEKIFQLKA 7589
QY 438 -----PTYVNLTP----- 445
Db 7590 IPKKKVPENPQVPEKVELTPLKVPGEKKVKRKLPERKPEKBEVVLKSVLRKRPEBEP 7649
QY 446 -----PEHRAVNL----- 454
Db 7650 KVPEKKLEKVKKPAVPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBP 7709
QY 455 ----- 454
Db 7710 PKPEAEVKTIKPPVPEPTPIAAPVTVPVVGKAAKAPKEAAKPKPIKVPKKTPTS 7769
QY 455 ----- 454
Db 7770 PIEAERRKLRPGSGGKPPDEAPPTYQLKAVPLKFVKKEIKDIILTESEFVGSSAIF 7829
QY 455 STSNLSLWNLQ 465
Db 7830 SPSTAITTMWK 7840
```

## RESULT 13

S20901

titin - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C&gt;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004

C:Accession: S20901; I46520

R:Label: S.; Gautel, M.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A:Title: Towards a molecular understanding of titin.

A:Reference number: S20997; MUID:92258380; PMID:1582406

A:Accession: S20901

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-6805 &lt;LAB&gt;

A:Cross-references: UNIPROT:Q28733; EMBL:X64696

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

R:Label: S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U.;

Nature 345, 273-276, 1990

A:Title: A regular pattern of two types of 100-residue motif in the sequence of titin.

A:Reference number: I46520; MUID:90238553; PMID:2129545

A:Accession: I46520

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 4235-5250 &lt;LA2&gt;

A:Cross-references: EMBL:X17329; NID:gl756; PIDN:CAA35207.1; PID:g930251

C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro

C:Keywords: muscle

## Query Match

Best Local Similarity 19.1%; Score 469; DB 2; Length 6805;

Matches 221; Conservative 84; Mismatches 155; Indels 6096; Gaps 72;

```
QY 4 KEMVVR-DVPMFVLISIFLL----- 24
Db 52 EEVVKEDLQKPLVDLKLGLTVKAGETIRLEAGVRGKPPPEVVTWKDQATDTRSPR 111
QY 25 -----VS 26
Db 112 AKIDTSADSKFSLTKAKRSDGGKVVVVTATNTAGSFWAYATVNVLDKPGVNRNLKIPDVS 171
QY 27 FINC----- 30
Db 172 SDRCTIRWDPPDDGGCEIQNYILEKCESKRWVWSTYSATVLTPTGTVTRLIEGNEYIFR 231
QY 31 -----KVMKALYNRP----- 41
```

Db 232 VRAENKIGTPPTESKPVIAKTYDRPGRDPDPPEVTKVSKBEMTVVMSPEYDGGKSITG 291  
QY 42 -----|:|:|:|----- 41  
Db 292 YLEKKEKHSVRWVPVNSAIPERRLKVNOLIPGHEYQFRVKAENEIGVGEPSLPSRPV 351  
QY 42 -----|:|:|:|----- 41  
Db 352 AKDPIPPGPINLKVVDTTKSSITLSWGRPVYDGGAPIIGYVVEVRPKIADASPDEGWK 411  
QY 42 -----|:|:|:|----- 41  
Db 412 RCNAAQLVRTEFTVTSIDENOEYFRVCAQNOVGIGRPAELKEAIKPKELPPEIDLD 471  
QY 42 -----WRLVLSKI----- 50  
Db 472 ASMRKLVVRAGCPIRLFAIVRGRPAKVTWRKVGIDNVVRKQGVLDVMTAFLVFNST 531  
QY 51 ----GKYL-----DQLKLEIL----- 63  
Db 532 RDSGKYSLTMNPAGEKAVFNVVRVLDTPGPVSDLKVSDVTYKTSCHVSWAPPENDGGSQ 591  
QY 64 -----|:|:|:|----- 63  
Db 592 VTHYVEKRDARTWSIVNPEVKTSQVTLNVPNGEYFVRVTA VNEYGPGVPADVPR 651  
QY 64 -----ROLETTISTK-----YNSKOP----- 80  
Db 652 VLASDGLSEDDPRKLEVTMTKNSATLAWLPRLDGGAKIDGYIISYREEDQPADRWTE 711  
QY 81 ----VKNLTM-----NTEFPQ----- 92  
Db 712 YSVVDLSLVITGLKEGKYKFRVAARNAVGVSLPREABGVYEAKEQLIPPKILMPEQIT 771  
QY 93 -----|:|:|:|-----XY 94  
Db 772 IKAGKLRLEAHVYKGPQPCIKWKGEDDVTSSHVAHKAESSILIKDVTBKSGY 831  
QY 95 ILA-----GP-----IONYSI 105  
Db 832 SLTAENSSGTDQIKIVMDRPGPPQPPFDISDIDADACSLSWHIIPLDGGSNITNIV 891  
QY 106 -----|:|:|:|----- 105  
Db 892 EKCDVSRGDVWTALASVTKTSRICGLIPGOBYFRVRAENRFGISEPLQSPKMLAQPPF 951  
QY 106 -----|:|:|:|-----TYLWDFYSTQLR-- 118  
Db 952 GVPSEPKNARVTKVANKDCIFVANDRPDSGGSPITGYLIERKGRNSLLWVKANDTAVRST 1011  
QY 119 -----KPAKYVYS----- 126  
Db 1012 EYPCAGLEGLYSFRIYALNKAAGSPSPKTEYVARTPDPVDPKBEVIDVTKSTVSLI 1071  
QY 127 -----|:|:|:|-----QYNHTAK 133  
Db 1072 WARPKHGSGSKIIGYFVEACKLPGDKWVRCNTTTHQIPHEEYTVTGLEENAQYQFRAIAK 1131  
QY 134 T-----|:|:|:|----- 134  
Db 1132 TAVNISOPSELRTPVTIHAENVPRIDLSVAMKSLLTVKAGTNVCLDATVPGKMPVSW 1191  
QY 135 -----|:|:|:|----- 134  
Db 1192 KKEGTVLKPAGIKWAMQORNICTELFVNRKOSGDYVITTAENSSGKSATIKLVLDPR 1251  
QY 135 -----ITFRPP----- 140  
Db 1252 GPPASVKINKMYSDRAMLSWEPLEDGGSBITNIVDKRETSSRNAQVSAVNPITSV 1311  
QY 141 -----PCGR----- 144  
|:|:|:|

Db 1312 EKLEIEGHEYQFRAICAENKYGVGDVPVTEPAIAKNPYDPPGRCDPVISNVTKDHMTYSWK 1371  
QY 145 -----|:|:|:|----- 144  
Db 1372 PPADGGSPITGYLLEKRETHAVNWKVRKPVIERTIKATGLOEGTEYEFVFTAINKAG 1431  
QY 145 -----|:|:|:|----- 144  
Db 1432 PGKPSDASKAVVAQDPLYPGPAPPKYVDITRSSVLSWCKPAYDGGSPILGVLVEVR 1491  
QY 145 -----VPSMTCLSEML- 155  
Db 1492 ADTNWVRCNLPQKLOKTRFVETGLMENTEYQFRVAVANKVGYSDPDPDGHCKPKDILI 1551  
QY 156 -----NYSKRDNG----- 164  
Db 1552 PPEGELDADLRKTLILBRAGVTMRLYVPKGRPPPKITWSPKPNVNLRIERIGLDIKSTDFDT 1611  
QY 165 -----BOGCGN-----FTTFNP 176  
Db 1612 FLRCENVKYDAGKYILTLSENSCKKGYTIWKVLDTPGPPVNVTVKEISRDSAYITWDP 1671  
QY 177 -----MFFNV----- 181  
Db 1672 PIVDGGSPIINYVVEKDAERKSWSTVTTECSKTSFRVSNLEEGKSYFRVVAENBYGIG 1731  
QY 182 -----PR----- 183  
Db 1732 DPGETRDAVKASETPPGVVDLVKLVTKSSCNIGWKPKRSDGSRITGVVDFLTENKW 1791  
QY 184 -----|:|:|:|----- 183  
Db 1792 QRWKSLSLOYSTKDLNEGQYTFRVSAENENGEGTSPBITVAVKDDVAVDLDLKDLP 1851  
QY 184 -----W-----NTKLYV----- 190  
Db 1852 LCVLAKENSFRLKI PHQKPAVPTWKKGEDPLATDTRSVSESSAVNTLVVYDCQKSD 1911  
QY 191 -----GPTK----- 194  
Db 1912 AGKYTITLKNVAGTKBGTLSIKVGCKPGIPTGPIKPFDEVTAEBATLKWGPPKDDGSEIT 1971  
QY 195 -----|:|:|:|----- 194  
Db 1972 NYILEKRDVNKNWVTCASAVOKTIFRVTRLHEGMEYTFRVSAENKYGVGEGLKSEPIVA 2031  
QY 195 -----|:|:|:|----- 194  
Db 2032 KHPFDVPDAPPPNIVDVHDSVSLTWDPRKTGGSPITGVHIEPKERNSLMKBEANKTP 2091  
QY 195 -----VNVDST 201  
Db 2092 IRMKDFKVTGLTEGLEEYFRVAINLAGVKPSLPSEPVVALDPIDPPGKBEVINVRNS 2151  
QY 202 IYFL-----|:|:|:|----- 205  
Db 2152 VTLIWPYKDGKHLTGVIVEKRDLPSTWKNKANHINVPDCAFTVDTOLVEGKYSEFR 2211  
QY 206 -----GLT----- 208  
Db 2212 AKNTAGAISAPSESTGTIICKDEYEAPTIVLDPITKDGLTIKAGDTIVLNAISILGKPLP 2271  
QY 209 -----ALLRYAOR----- 217  
Db 2272 KSSWSKAGDIRPSDITQITSTPTSSMLTVKYASRKDAGEYTTATNPFPGTKEHVRVTV 2331  
QY 218 -----|:|:|:|-----NC 219  
Db 2332 LDVPFPGPIBISNVSAEATLTWTPLEDDGSPKSVVLEKRETSRLLTWVAEDIQSC 2391  
QY 220 TH-----|:|:|:|----- 221  
Db 2392 RHVTKLIQNEYLFRVSAVNNHYGKBPVQSEPVKMDVDRFPFGPPGKBEVSNVTQNTAT 2451  
|:|:|:|



QY 222 ----- 221  
Db 2452 VSKRPTDDGSEITGYVERREKGLWRATKTPVSDLRCKVTGLQEGNTYFRVSAE 2511  
QY 222 ----- 221  
Db 2512 NRAGIGPPSDASNYVLMKDVAYPGPPSNARVTDTTKKSASLANCKPHYDGLLEITGVV 2571  
QY 222 ----- 221  
Db 2572 EHQKVGDETWKDTGTPALRITEFVVPDLTKKYNFRISAINDAGVCEPAVDPVEIVE 2631  
QY 222 -----SP 223  
Db 2632 REMAPDFELDAELRRLTVRAGLSIRIFVPIKGRPAPEVTWKDINLKTRANIENTESP 2691  
QY 224 ----- 223  
Db 2692 TLLIPECNRYDTGKFVMTIENPAGKSGFVNVRLDTPGPVLNLRPTDITKDSVTLHWD 2751  
QY 224 -----LFRV----- 236  
Db 2752 LPLIDGSGRIITYIVEKREATRKSYSVTTKCHKCTYKVTGLSGCEYFFRVMAENEXGI 2811  
QY 237 -----PKYINGTK----- 244  
Db 2812 GEPSETKEPVKASEAPSPDSLINIMDITKSTVSLAWPKPKHGGSKITGYVIEAQRKGS 2871  
QY 245 ----- 244  
Db 2872 QMTHITTVKLECVVRNLTEGEYTFQVMVNASGRSAPRESRPVIVKBQTMPLPELDLGR 2931  
QY 245 -----LKNVNR----- 250  
Db 2932 IYQKLVIAKAGDNIKVEIPVLRPKPTVTWKGDQVLKQTORVNVENTATSTILNISECV 2991  
QY 251 ----- 250  
Db 2992 RSDSGPYLTAKNIVGEVDVITIQVHDPGPPTGPIKFDEVSDFVTFWSPEPPENDGV 3051  
QY 251 -----KLRQAOPVKEOPEKA----- 267  
Db 3052 PISNYVEMRQDSTTWVELATTVIRTQKATRLTTGVEYQFRVKAQNRGVGPGITSAS 3111  
QY 268 ----- 267  
Db 3112 IVANYPFKVPGPPTQVTAUTKDSMTISWHEPLSDGSGPILGYHVERKERNGLMQTVS 3171  
QY 268 ----- 267  
Db 3172 KALVPGNIFKSSGLTDGIAEPRVIAENWACKSPKSPBVLALDPIDPGKPIPLNIT 3231  
QY 268 ----- 267  
Db 3232 RHTVTLKWAPEYTGFGKITSYIVEKRLPNRMLKANFSNILENEFTVSLGTEDAAYEF 3291  
QY 268 ----- 267  
Db 3292 RVIAKNAAGATSPSPSPSDAITCRDDVEAPRILVDVRFKOTVILKAGEAFKLEADVSGRP 3351  
QY 268 ----- 267  
Db 3352 PPTWETKDGKEBGTGKLEIKIADFSTYLINKDSRRSDSGAYILTATDPGFAKHIFNV 3411  
QY 268 ----- 267  
Db 3412 KVLDRPGPEGLAVSEVTSEKCVLSWLPPLDDGGAKIEHYIVQKRETSRLATWNVASEV 3471  
QY 268 ----- 267  
Db 3472 QVTKLVTKLLKGNEYIFRVMAVNKYGVGEPLSEPVLANVPYGPDPKPNPVTITKD 3531

QY 268 -----KKT----- 270  
Db 3532 SMVVCWGHPSDDGSEIINYIVERRDKAGORWKNKKTVDLRFKVSGLTEGHSEYEPRI 3591  
QY 271 -----QSTTTPYF----- 278  
Db 3592 MAENAGISAPSRISPPYKACDAVPKPGPCNPRVLDTSRSSISIAWNKPIYDGGSEITG 3651  
QY 279 ----- 278  
Db 3652 YMBIALPEBDEWKIVTPPAGLKATSVITNLVENQBYKIRIYAMNSEGLGEPALVPOTP 3711  
QY 279 ----- 278  
Db 3712 KAEDRMLPPRIELDADLRKLWIRACCTLRFLVPIKGRPDEPVKWTREHGESLKDASIES 3771  
QY 279 --SVT-----TSAALNV----- 288  
Db 3772 TSSYTLIIGVNVRPDSGKYLITVENSNGSKSAFVNRVLDTPGPPQDLKVEKVTKTSVT 3831  
QY 289 ----- 288  
Db 3832 LTWDPPLLDGGSKIKNYIVEKRESTRKAYSTVATNCHKTSWKVDQLQEGSSYFRVLAEN 3891  
QY 289 ----- 288  
Db 3892 EYGILPAETAESVKASERPLPPGKITLVDVTRNSVLSWEKPEHGGSRILGYIVEMQS 3951  
QY 289 -----TNNVT-----YSITTA----- 300  
Db 3952 KGSQWATCATVKTETATITGLIOGBEYSFRVSAQNEKIGSDPRQLSPVIAKOLVPPA 4011  
QY 301 ----- 300  
Db 4012 FKLLFNTFTVLAGEBKLIDVPPFGRPTPTVTHKDDVPLKQTRVNAESTENSILLSIKE 4071  
QY 301 ----- 300  
Db 4072 ACREDVHYVVKLSNSAGEATELNAIILDKPGPTGPFVKMDEVTAADSIISWBPVKYDG 4131  
QY 301 -----RRVSTST----- 307  
Db 4132 GSSNNVIVEKRTDSTTTTQIVSATVARTTIKASRLKTGEYQPRIAENRYKSTYLSNS 4191  
QY 308 -----IAYR----- 311  
Db 4192 EPIVAQYFVKVPGPGTPTFTLSSRDSMEVQWNEPVDNGGSRVIGYHLERKERNLSILWVK 4251  
QY 312 ----- 311  
Db 4252 LNKTPIPQTKFKTTGLEEGIEYFRVSAENIVGIGKPSKVSECVYVARPCDPPGRPERII 4311  
QY 312 -----PDSSFMK-----SIMATQLR----- 326  
Db 4312 VTRNSVTLOWKXPTYDGGSKIYGVVEKBLPDGRWMAKSFNTIMDTQFEVTGLVEDHRY 4371  
QY 327 ----- 326  
Db 4372 EPRVIARNAAGVSESPSESTGAITARDEIDPPRISMDPKYKDTIVVHAGESFRIDADIYG 4431  
QY 327 -----DLAT----- 330  
Db 4432 KPIPTQWIKQDOELSNTARLEIKSTDFATSLSVKDAFRVDSGNVYLKAQNVAGERSVTV 4491  
QY 331 -----W-----VYTL-- 336  
Db 4492 NVKVLDRPGPEGPVIVSGVTAEKCTLAWKPLQDGGSDIINYIVERRETSRLVMTVDA 4551  
QY 337 -----RY-----RONPF----- 343  
Db 4552 NVQTLSCVKTKLEGNEYIFRVMAVNKYGVGEPLSEPVIAKPNFVVPDAPAPCECTVT 4611  
QY 344 ----- 343

Db 4612 KDSMVVWVWPASDGGSEILGYVLEKRDKEGIRWTRCHKELIGELRLRVTLGIENHNYEF 4671  
QY 344 -----CEP----- 346  
Db 4672 RVSAENAGLSBSPSPAYQKACDPYKPGPPNNPKVMDITRSSVFLSKPIYDGGCEI 4731  
QY 347 ----- 346  
Db 4732 QGVIVEKDVSGEWTMCTPPTGINKNTINEVEKLEKHEYNFRICAVNKAGVDHADVEG 4791  
QY 347 ----- 346  
Db 4792 PVIVEKLEAPDIDLLELRKIINIRAGGSLRFLVPKIGRPTPEVKWGVGDGEIRDAII 4851  
QY 347 -----SRNR----- 350  
Db 4852 DSTSFTSLVDNVNRYDSKTYLTLENSSGTKSAFVTVRVLDPSPPPVNLKVTETIKDS 4911  
QY 351 -----TA 352  
Db 4912 VSIWEPPLDGGSKIKNYIVEKRDSTRKSYAAVTVNCHKSSWKIDQLQEGCSYFRVTA 4971  
QY 353 VSEF----- 356  
Db 4972 ENSYGIGLPARTADPIKVAEVPQPGKITVDDVTRNSVLSWTKPEHDGSGKIIQIVIVEM 5031  
QY 357 -----MKNTHVLRN----- 366  
Db 5032 QAKHSEKNECARVKSLEAVITNLTOGSEYLFRRVAVNKEGRSDPRSLAVPIVAKDLVIE 5091  
QY 367 -----ETPYT-----IYGLDMSSLYY 383  
Db 5092 PDVKPAFSSYVQGDLEIEPISGRPKPTITWKGGLPLKQTRINVAUSDLTITLSI 5151  
QY 384 NET-----MFVENTAS----- 395  
Db 5152 KETHKDDSGHYGITVANVVGQKTASIEIITLDPDPKPGPKFDEVSASISITLSNPNPLY 5211  
QY 396 -----DSNKTPT----- 403  
Db 5212 TGGCQITNVIIVKRDITTTTWDVVSATVARTTLKTKTGTEYQFRIPPENRVGQSAL 5271  
QY 404 -----SPSMGF-----ORTFIDP 416  
Db 5272 DSEPIVAQYPYKEPGPGTPTVATSKDSVVQWHEPINNGSPITIGYHLERKERNI-- 5329  
QY 417 LWDYLDLSLLFLD----- 428  
Db 5330 LMTKVDKSIHDTQKALNLEBIEYFRVYAENIVGVGKASKNSECYVARDPCDPPGTP 5389  
QY 429 ----- 428  
Db 5390 EAIIVKRNEITLQWTKPVYDGGSMITGYIVEKRDLPGRWMAKSFNTVIEQTVTSGLTE 5449  
QY 429 ----- 428  
Db 5450 DQRYEFRVIAKNAAGTMSKPSDSTGPIAKDEVELPRISMDPKFRDITIVNAGETFRLEA 5509  
QY 429 -----EIRN-----FSLRS----- 437  
Db 5510 DVHGKPLTIEWLRGDKEVESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNAVASK 5569  
QY 438 -----PTTV----- 441  
Db 5570 SFPVNVKLDPPGPEGPVQVTGTCCKTLTWSPLQDGGSDIPHYVVEKRETSRLAWT 5629  
QY 442 ----- 441  
Db 5630 VVASEWVNSLKITKLEGNBYIFRIMAVNKYGVGEPLASAPVLKMNPFVVPKPSLEV 5689  
QY 442 ----- 441

Db 5690 TNIAKDSMTVCWNRDSDGGSEITGYIVEKDRSGIRWIKCNKRVRTDLRFRVTGLTEDH 5749  
QY 442 ----- 441  
Db 5750 EYEFVRSAENAGVGEPSPATVYVKACDPVFKPGPTNAHVVDTTKNSITILANGKPIYDG 5809  
QY 442 ----- 441  
Db 5810 GSEVLGYIIEICKADEEEMQIVTPQTGLKANRFBISKLIEHQYKIRVCALNKVGLGEAA 5869  
QY 442 ----- 441  
Db 5870 SVPGTVPKEDKLEAPELDLDELKGIIVRAGGSARIHIPKGRPTDITWRSREGEFTD 5929  
QY 442 ----- 441  
Db 5930 KVQVEKGVNFTQLSIDNCNRDAGKIYKLENSSGTKTAFVTVKVLDTPGPPQNLAKEY 5989  
QY 442 -----NLT----- 444  
Db 5990 KDSAVLWPEPPIIDGGAKVRNVYIDKRESTRKAYANVSSKNKTTFKVENLTGAIYYF 6049  
QY 445 -----PPEH----- 448  
Db 6050 RVMAENEFVGVPVETVDVAKAAEPSPPGKVTLTDVTSQTSASLWKEKPEHDGSGRVLYG 6109  
QY 449 ----- 448  
Db 6110 VVEMQPKTEKWSVAVBSKVCNAVVTGLSSGHEYQFRVKAYNEKSGSDPRVLGVPIAKD 6169  
QY 449 -----RAAV----- 452  
Db 6170 LTQPSKLPFKRYSVQAGEDLKIEIPVIGRPRPEIPFWKDGSEPLRQTTTRVNVEETATST 6229  
QY 453 -----NLT----- 456  
Db 6230 ILHIKSSKDDFGKYTITATNSAGTATENLSVIVLEKPGPPGVPRFDEISADFFVLSWE 6289  
QY 457 ----- 456  
Db 6290 PPAYTGGCQISNVIYVEKRDITTTTWHIVSATVARTTIKTKLKTGSEYQPRIYAENRYGK 6349  
QY 457 -----SN 458  
Db 6350 STSLDSKPVIVQYFPFKEPGPGTPTVTSVSRDQMLVQWHEPVDNDGGSKVLGYHLEOKEN 6409  
QY 459 SLMW----- 462  
Db 6410 SILWVKNKTLIQDTKFKTGTGLDEGLEBYFKVSAENIVGLASLAKCPNAPVARDPCDPPG 6469  
QY 463 ----- 462  
Db 6470 RPEAIVITRNNVTLKWKPAYDGGSKIYGYVEKKDLPGRWMAKSFNTVLETFETVSG 6529  
QY 463 ----- 462  
Db 6530 VEDQRYEFRVIRNAAGNLSEPSSESGAITARDEIDAPNASLDPKYKDVIIVHAGETFFVL 6589  
QY 463 -----WLQ 465  
Db 6590 EADIRKGPIDVVMWLK 6605

## RESULT 14

A88852

protein unc-22 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C;Accession: A88852

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A;Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_elegans/ and  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: A88852  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-6831 <STO>  
A:Cross-references: UNIPROT:Q23550; GB:chr\_IV; PIDN:CAA98081.1; PID:G3881830; GSPDB:GN00  
C:Genetics:  
A:Gene: unc-22  
A:Map position: 4  
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 19.1%; Score 469; DB 2; Length 6831;  
Best Local Similarity 3.5%; Pred. No. 2.4;  
Matches 220; Conservative 92; Mismatches 131; Indels 5900; Gaps 75;

Qy	2	GRKEMNR-----	9
Db	451	GOSSNVKIBQSDVZEELMKHRDAEDYQKEQSKQLQAEYKRVARRSKSKSPAP	510
Qy	10	-----	9
Db	511	QAKSTTSSEGRQAEVEHKRSSVRPDDEESQLDEIPSSGLTIPERRRELLGQVGE	570
Qy	10	-----DVPKMFVLISFLVS-----	26
Db	571	SDDEVSEISLPSFAGGPKRKTDKPK-----KVSIAPIVSTNKSDDPSTPRRSSI	625
Qy	27	-----FINCKVMS-----	34
Db	626	DMRESVQELIKTSTPLVPSGASGAPKIVEPENVTVENETAILTCVSGSPAPTR	685
Qy	35	-----	34
Db	686	WFKGSREVISGRPKHITDGKEHTVALLLKCRSQDEGPTLTITENVHGTDSADVCLLV	745
Qy	35	-----	34
Db	746	SDNGLDFRAMLKHSQAFQDGGGGGGGEGKEKPMTEARRQSLLFPQKVEKWDIP	805
Qy	35	-----KALYNRP-----	41
Db	806	LPEKTVOQVDKICEWKTCTSRPNKIRWYKDRKEIPSGGLKYKIVIEKNVCTLIINPE	865
Qy	42	-----	41
Db	866	VDDTGKTCBANGVPTHAQLTVLEPPMKYSFLNPLPNTQEIYRTKQAVLTCKVNTPRAPL	925
Qy	42	-----	41
Db	926	VYRGSKAIQEGDPRFIEKDAVGRCTLTIKEVEEDQAEWTARITQDVFSKVQVYVEEP	985
Qy	42	-----W-----	46
Db	986	RHTFVPMKSQKNESDLATLETVDNDKDAEYVWVHDKRIDIDGVKPKVSSNRKRLI	1045
Qy	47	LS-----	48
Db	1046	INGARIEDHGEYKCTTKDDRTMAQLIVDAKNKPIVALKDTEVIEKDDVTLACQTKDTP	1105
Qy	49	-----KIGK-----YKLDQL-----	58
Db	1106	GIWFRNGKQISSMPGKPGFQTSRNGHTLKIGKIEKNADYVEIDQAGLRGCNVTVLEA	1165
Qy	59	-----KLEILR-----	64
Db	1166	EKRPILNWPKKIEAKAGEPCVVKVPFIQKTRRGDPKQAQLKNGKPIDEMRKLEVEII	1225
Qy	65	-----QLETT-----ISTKYNVSKQ-----	79
Db	1226	KDDVAEIVPNPQLADTKWALELGNAGTALAPPELFDKDKPKPKGPLETK-NVTAEG	1284
Qy	80	-----PVK-----NLTNTEF-----	90

Db	1285	LDLVNGTDPDPDEGAPVAYIIMQGRSGNNAKVGETKGTDFKVKDLKEHGEYKFRVKAL	1344
Qy	91	-----	90
Db	1345	NECGLSDPLTGESVLAKNPYGVPGKPKMDAIDVDKDHCTLAWEPPESDGGAPITGYIE	1404
Qy	91	-----	90
Db	1405	RREKSEKMDHQVQTKPDCCCLTDKVVVEDKEYLYRVKAVNKGAGPGDPCDHGKPIKMAK	1464
Qy	91	---POY-----	93
Db	1465	KASPEFTGGGIKDLRLKVGETIKYDVPISGBPLPECLWVNGKPLKAVGRVMSERGH	1524
Qy	94	-----	93
Db	1525	IMKIENAVRADSGKFTITLKNSSGCDSTATVTVVGRTPPKGPLDIADVCADGATLSWN	1584
Qy	94	-----YILAG-----	98
Db	1585	PPDDGGDPLTGYIYEAQMDNKGKIEVGKVDPTNTTLKVNGLRNKGYKFRKAVNNE	1644
Qy	99	-----PIQNY-----	103
Db	1645	GESEPLSADQYTIKDPWDEPGKPRBEITFDADRIDIAWEPHKGAGPIEYIVR	1704
Qy	104	-----	103
Db	1705	DPDTKEWKEVVPDNTNASISGLKEGKEYQFRVAVNKGAGPGQSEPESEKQAKPKPIPA	1764
Qy	104	-----SIT-----YLMF-----	110
Db	1765	WLKHDNLKSIIVKAGATVRWEVKIGGEPPEVKFKGNOQLENGILQTLIDTRKNEHTLC	1824
Qy	111	-----	110
Db	1825	IPSAMRSDVGEYRLTVKNHSHGADBEKANLTVLDRPSKPNGLPVSDFEDNLNLSWKPPD	1884
Qy	111	-----DYSTQ-----LRPKAYVY-----	125
Db	1885	DDGGEPIEYVEVEKLDATGRWVPCAKVKDTKAHIDGLKGGQTYQFRVAVNKEGASDAL	1944
Qy	126	-----SQY-----	128
Db	1945	STDKDKAKNPYDEPGKTGTDPVVDADRVSLWEPEPKSGDGPITQYVIEKKKGHGRD	2004
Qy	129	-----	129
Db	2005	WQCGKVSQDQTNABILGLKEGEEYQFRVAVNKGAGPGAEASDPSRKVVAKPRNLKPWIDR	2064
Qy	130	HTAKTITFR-----PP-----	140
Db	2065	EMKTIITIKGNDVEFDVVRGEPPEPKKEWIFNEKPVDDQKIRIESEDKYTRFVLRGATR	2124
Qy	141	-----	140
Db	2125	KHAGLYTLTATNAGSKSHSEVIVILGKPSPLGPLEVSNVVEDRADLEWKVPEDDGGAP	2184
Qy	141	-----PCGR-----	144
Db	2185	IDHYEIKMDLATGRWVYPCGRSETTKTTPVNLQPGHEYKFRVAVNKEGESDPLTTNTAI	2244
Qy	145	-----VPSMTCLEMLN-----	156
Db	2245	LAKNPYVPGKVPKPELVVDMDKDHVDLAWNAPDDGGAPIEAFVIEKKDKNGRWEALVVP	2304
Qy	157	-----VSKENDTG-----	164
Db	2305	GDQKTATVPNLKEGEEYQFRISARNKAGTGDPSDPSRVVAKPRNLAPRIHREDLSDTTV	2364
Qy	165	-----EQCGGNFT-----	172

Db 2365 KVGATLKFIVHIDGEPADVTWSENGKIGESKAQIENEPYISRFALPKALRKQSGKYI 2424  
QY 173 -----TEN-----PM-FFNVPR 183  
Db 2425 TATNINGTSDVTIINKVSKETPKGPTEVTVDFEDRATLDWKPPEDDGGEPFIEFYIEK 2484  
QY 184 WNTK--LVY-----GPTKVNVD5-----199  
Db 2485 MNTKGIWVPCGRSGDTHFTVDSLNGDHYKFRKAVNSGSPDPLETETDILAKPPDR 2544  
QY 200 -----QTIYF 204  
Db 2545 PDRGRPEPTDWDSDHVDLKWDPPLSDGAPIEBYQIEKTKYGRWEPALTVPGGOT---2601  
QY 205 LGLTA-----209  
Db 2602 ---TATVPDLTPNEEYFRVAVNVKGGSPDSASKAVIAKPNLKHIDRDLAKNLTIK 2658  
QY 210 -----LLRYAQRNCTHSP- 223  
Db 2659 AGOSISFDPVPSGEPAPTWTWHWPDNREIRNGGRVKLDNPEYQSKLVVKMERGDSGT 2718  
QY 224 -----223  
Db 2719 IKAVNAGEDATKINVIDKPTSPNGPLDVSDVHGDHVTLNWRAPDDGGPIENYIE 2778  
QY 224 -----223  
Db 2779 KYDTASGRWVPAKVAGDKTTAVVDGLIPCHEYKFRVAANAEGSDPLETGTTLAKDP 2838  
QY 224 -----223  
Db 2839 FDKPGKTNAPEITDMDKHVDLEWKPPANDGGAPIEYVYVEMKDBSPFWDVAHVPAQ 2898  
QY 224 -----223  
Db 2899 TNATVGNLKEGKYEFIRAKNKAGLGDPSDSASAVAKARNVPPVIDRNSQIEIKVKAG 2958  
QY 224 -----YLVNAM 229  
Db 2959 DFLSNIPVSGEPTTITWTFTGTPVESDDRMKLNNEDGKTKPHVKRLRSDTGTIYIAE 3018  
QY 230 SRN-----232  
Db 3019 NENGTTAEVKVTVLDHPSSPRGPLDVTVNIVKDCOLAWKPEDDGGAEISHYVIEKQDA 3078  
QY 233 -----232  
Db 3079 ATGRWTACGESKDTNFHVDLDTQGEYKFRKAVNRHGDSDPLEAREALIAKDPDRADK 3138  
QY 233 -----232  
Db 3139 PGTPEIVDMDKHADLKWTPPADDGGAPIEGYLVEMRTPSGDWVPAVTVGAGELTATVDG 3198  
QY 233 -----LFRV-----PKVI-----240  
Db 3199 LKPGQTYQFRVKALNKAGESTPSRFTWAKPHLAPKINRDMFVAQRVAGQTLNFDV 3258  
QY 241 -----NKTML--KNTMR-----250  
Db 3259 NVEGEPAPKIEWFLNGSPSSGGNTHIDNNNTDNNKLTSTKSTARADSGKYIVATNESGK 3318  
QY 251 -----250  
Db 3319 DEHEVDVNILDIPGAPEGLRHKOITKESVVLKWDEPLDDGGSPITVYVVEKQEDGRWV 3378  
QY 251 -----KLK-----253  
Db 3379 PCGETSDTSLKVNKLSEGEYKFRKAVNRQGTSAPLTSDHAIKVNPPDEPDPTDTP 3438  
QY 254 -----253  
Db 3439 VDWKDHVDLEWKPPANDGGAPIDAYIVEKKDKFGDWVECARVDGKTKTATADNLTPGET 3498

QY 254 -----RKOAP-----258  
Db 3499 YQFRVKAVNKAGPGKPSDPTGNVNAKPRMAPKLNLAGLLDLRIKAGTPIKLDIAFEGEP 3558  
QY 259 -----258  
Db 3559 APVAKKANDATIDTARADVTNTPTSSAIHIFSAVRGDTGVYKIIIVENEHKGDTAQCNV 3618  
QY 259 -----258  
Db 3619 TVLDVPGTPEGPLKIDEIHKEGCTLNWKPTDNGDVLHYIVEKMDTSGRTGWEVGTFF 3678  
QY 259 -----VKEQF-----263  
Db 3679 DCTAKVNKLVPGEYAFRVKAVNLQGESKPLEAESEPIIAKNQFVDPDPVKPEVTDWDKD 3738  
QY 264 -----EKKAK--TOSTTTPYFSYTTSAALNVTNTVY--294  
Db 3739 RIDIKWNPNTANNGAPVTGYIVEKKEGSAIWTGKTP--GTTFSADNLKPGVEYEF 3795  
QY 295 -----SITTAARRV-----303  
Db 3796 VIAVNAAGSPSPDPTDQITKARYLKPILTSRKIKIKAGFTHNLEVDPIGAPDPTAT 3855  
QY 304 -----STSTIAY-----310  
Db 3856 WTVGDSGAALAPELLVDAKSTTIFPFSADNLKPGVEYEF 3915  
QY 311 -----310  
Db 3916 PSAPEGLEVSDVTKSCVLNWKPPKDDGGAESYVVEKRDTKTNTWVPSAFVTGTSI 3975  
QY 311 -----310  
Db 3976 TVPKLTGEGHEYFRVMAENTFGRSDSLNTDPEVLAKDPFGTGPGRPEIVDTDNHIDI 4035  
QY 311 -----310  
Db 4036 KNDPPRDNGGSPVDHYDIERKDAKTGRWIKVNTSPVGTAFSDTRVQKGTTEYRVAVN 4095  
QY 311 -----RPDSSEPMK 318  
Db 4096 KAGPGQPSDSSAAATAKPMHEAPKFDLDDGKEFRKAGEPLVITPTFTASQPDLSWTK 4155  
QY 319 S-----319  
Db 4156 EGGKPLAGVETTDTSOTKLVIPSTRSDSGFVKIKAVNPYGEAEANIKITVIDKPGAPENI 4215  
QY 320 -----319  
Db 4216 TYPVSRHTCTLNWDAPKDDGGAETAGYKIEYQEVGSQIWDKVPGLISGTAYTVRGLBHG 4275  
QY 320 -----319  
Db 4276 QOYFRFIRAENAVGLSDYCGQVVPVVKDPDPDPGAPSTPEITGYDTNQVSLAWNPRDDG 4335  
QY 320 -----IMATO-----324  
Db 4336 GSPILGYVVERFEKRGGDWAPVKMPVKGTECIVPGLHENETYQFRVAVNAAGHGEPS 4395  
QY 325 -----324  
Db 4396 NGSEPVTCRPYVEKPGAPDAPRVGKITKNSAELTWNRLRDGGAPIDGYIVEKKKLGDN 4455  
QY 325 -----LRDLA-----329  
Db 4456 WTRCNDKVPDRTAFEVKNLGEKEEYFRVIAVNSAGEGEPKPSDLVLIBEQGRPIFDI 4515  
QY 330 -----329  
Db 4516 NNLKDIITVRAGETIQIRIPYAGGNPKPIIDLFGNGSPIFENERIVVDNPGIEVITTTGS 4575

QY 330 -----TW----- 331  
Db 4576 KRSAGPYKISATNKYKGDCKLVNVLDPAGKPTGPIRATDIOADAMTWSRPPKONGG 4635  
QY 332 -----VYTLIRY----- 339  
Db 4636 DAIYVVEKRTPGGDWTVGHPVGTTLRVRNLDANTPYEPRVRAENQYGVGEPLTDDA 4695  
QY 340 -----QNPFCPE----- 346  
Db 4696 IVAKNPFDPGAPGQPEAVETSEBALTQWTRPTSDGAPIQGVVIEKREVGSTEWTKAA 4755  
QY 347 -----SENRTA----- 352  
Db 4756 FGNILDTKHRVTGLTPKTYEPRVAAYNAAGGEYSVNSVPITADNAPTRPKINNGMLTR 4815  
QY 353 ----- 352  
Db 4816 DILAYAGERAKILVPPAASPAPKVTFSGENKISFTDPRVKVYSDPLATITIEKSELTD 4875  
QY 353 ----- 352  
Db 4876 GGLYFVELENSQGSASIRLKWVDKPKASPOHIRVEDIAPDCCTLYWMPSSDGSPTN 4935  
QY 353 -----VSEFMKTHVL-----IRNETPYTI----- 372  
Db 4936 YIVEKLDLRHSDGKWEKVSFVRNLNTYVGGLIKDNRVFRVRAETQYGVSEPCELADW 4995  
QY 373 ----- 372  
Db 4996 VAKYQFVNPQPEAPTVRDKDSTWAELEWDPDRGGSKIICYQVQYRDTSSGRWINAKWD 5055  
QY 373 ----- 372  
Db 5056 LSEQCHARVTLRQNGEPEPRIIAKNAAGSPKSPSPSERCQKSRFGPGPPPIHVGAISI 5115  
QY 373 -----YG----- 374  
Db 5116 GRNHCTITWMAPLEDGGSKITGYNVEIREYGSTLTWASDVNVRPEFTVDKREFNDYE 5175  
QY 375 ----- 374  
Db 5176 PRVAINAGKIGISLPSGPIKIQESGSRPQIVVKPEDTAQPNRRRAVTFCEAVGRPEP 5235  
QY 375 -----TLMSSLYNETMFV----- 389  
Db 5236 TARWLNRGRELPESSRYRFEASDGYKFTIKEVWDIDAGEYTVESNPNYSGSDTATANLV 5295  
QY 390 ----- 389  
Db 5296 QAPVIEKDVNTILPSGDLVRLKIYFSGTAPFRHSLVLRNEEDMDHPTIRIVEFDDHI 5355  
QY 390 -----ENKTASDSNKT----- 401  
Db 5356 LITIPALSVREAGRYEYTVSNDSEAGTGFMLNVTGLPEAPQGPLHISNIGPSTATLSWR 5415  
QY 402 ----- 401  
Db 5416 PPVTDGGSKITSYVVEKEDLSKDEWTVTSNVKDMYIVTGLFENHEVEYFVRSQAENGI 5475  
QY 402 -----PTSP-----SMGFORTFIDPLWD----- 419  
Db 5476 GAPLVSEHPPIARLPFPDPPTSPLEIVQVGGDVYTLWSQR-----PLSDGGGRLRGYIVE 5531  
QY 420 -----YLDLILF----- 426  
Db 5532 KQEBEHDEWFRCNQPNPPNNYVNLIDGRKYRYRVPFVANDAGLSDLAELDQTLFOASG 5591  
QY 427 ----- 426  
Db 5592 SGECPKIVSPLSDLNBEVGRCVTFECEISGSPREYRWFKCKELVDTSKYTLINKGDQK 5651  
QY 427 ----- 426

Db 5652 VLIINDLTSDDADEYTCRATNSSGSTRANLRIKTKPRVFIPPKHYGGYEAQKGETIEL 5711  
QY 427 ----- 426  
Db 5712 KIPYKAYPQGEARWTKDGEKIENNSKFSITTTDDKPFATLIRISNASREDYGEYRVVVVSVG 5771  
QY 427 -----LDEI----- 430  
Db 5772 SDSGTVNVTVADVPEPPRFPPIIENILDEAVILSWKPPALDGGSLVTNYTIEKREAMGWS 5831  
QY 431 ----- 430  
Db 5832 SPCAKSRYTYTIEGLRAGKQYEFRIIAENKHGSKCPCEPTAPVLIPODERKRRRGYDVD 5891  
QY 431 ----- 430  
Db 5892 EQGKIVRGKGTSSNYVDYDIWKQYYPQVEIKHDHVLDOHYDIEBELGTAGFVWHRV 5951  
QY 431 -----RNFSS-----LRSPYVNL----- 443  
Db 5952 TERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAFEDDNEVMVMIYBFM 6011  
QY 444 ----- 443  
Db 6012 SGGELFEKVADEHNKMSDEBAVEYMRQVCKGLCHMHNHNNYVHLDLKPENIMPTTKRSNEL 6071  
QY 444 ----- 443  
Db 6072 KLIDPGLTAHLDPKQSVKVTGTAEFAAPEVAEKPVGYYTDMHSGVGLSYLLSGLSPF 6131  
QY 444 ----- 443  
Db 6132 GGENDETLRNVKSCDMNDDSAFSGISEGDKPIRKLLADPNTRMTIHOALEHPMLTP 6191  
QY 444 -----TPP-----BHR----- 449  
Db 6192 GNAPGRDSQIPSSRYTKIRDSIKTKYDAMPEPLPLGRISNYSLSLRKHPQEIYDAFW 6251  
QY 450 ----- 449  
Db 6252 DRSEAQRPFIVKPYGTEVGEGQSANFYCRVIASSPPVVVTHKDDRELKQSVKYMRYNGN 6311  
QY 450 -----RAVN----- 453  
Db 6312 DYGLTINRVKDDKGEYTVRAKNSYGTKEBIVFLNVTNRHSBPLKPEPLEPKKAPSPRV 6371  
QY 454 ----- 453  
Db 6372 BEFKERSAPFTFHLNRNLIQKNHQCKLTCSLQGNPNPTIEMKDGHPVDEDRVQVSFR 6431  
QY 454 ----- 453  
Db 6432 SGVCSLEIFNARVDDAGTYTATNDLGDVSECVLTVQTKGGEPIPRVSSFRPRRAYDT 6491  
QY 454 LST----- 456  
Db 6492 LSTGTOVERSHSYADMRRRSLIRDVSPDVSRAADDLTKITNELPSPFTAQLSDSETEVG 6551  
QY 457 ----- 456  
Db 6552 SAEFSAAVSQGPPELIEWLHNGERISESDSRFRASYVAGKATLIRISDAKKSDEGQYLCA 6611  
QY 457 SNS----- 459  
Db 6612 SNSAGQEQTRATLTVKGDQPLNQHACQAVESELURVTKHLGGEIVNNGESVTFEARVQGT 6671  
QY 460 -----LM----- 461  
Db 6672 PEEVLWNRNGOELTNGDKTISIQDGETLSFTINSADASDAGHYQLEVRKGTNLVSVASL 6731  
QY 462 -----WW 463  
|:

Db 6732 VVGEKADPPVTRLPSSVSAPLGSGSTAFTEPENVGLTVQWF 6774

RESULT 15

S57242

twitchin [similarity] - Caenorhabditis elegans

N:Alternate names: myosin-regulating protein

N:Contains: protein kinase (EC 2.7.1.1.)

C:Species: Caenorhabditis elegans

C>Date: 28-Oct-1995 #sequence revision 24-Oct-1997 #text change 09-Jul-2004

A:Accession: S57242; S07571; S06797; S57218; T27934; T28030

R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.

submitted to the EMBL Data Library, February 1993

A:Description: Additional sequence complexity within twitching of Caenorhabditis elegans

A:Reference number: S57242

A:Accession: S57242

A:Molecule type: DNA

A:Residues: 1-6839 <BEN1>

A:Cross-references: UNIPROT:Q23550; EMBL:L10351

A:Experimental source: var. Bristol

R:Benian, G.

submitted to the EMBL Data Library, November 1989

A:Reference number: S07571

A:Accession: S07571

A:Molecule type: DNA

A:Residues: 792-6839 <BEN2>

A:Cross-references: EMBL:X15423; NID:G6897; PIDN:CAA33463.1; PID:G6898

A:Experimental source: var. Bristol

R:Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.

Nature 342, 45-50, 1989

A:Title: Sequence of an unusually large protein implicated in regulation of myosin activity

A:Reference number: S06797; MUID:90044042; PMID:2812002

A:Accession: S06797

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 806-1175; 1178-1998, 'Y', 2000-3040, 'I', 3042-3335, 'I', 3337-5693; 5696-6359, 'I', 6361-6839

A:Cross-references: EMBL:X15423

A:Experimental source: var. Bristol

R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.

Genetics 134, 1097-1104, 1993

A:Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded protein

A:Reference number: S57218; MUID:93387664; PMID:8397135

A:Accession: S57218

A:Molecule type: DNA

A:Residues: 2-99; 108-194, 'Q', 196-206; 374-468; 658-753 <BEN4>

A:Experimental source: var. Bristol

R:White, S.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z20442

A:Accession: T27934

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 'MGIPGKKCKQ', 19-6839 <W1>

A:Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a

A:Experimental source: clone ZK617

R:Harris, B.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z20458

A:Accession: T28030

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 'MGIPGKKCKQ', 19-6839 <W2>

A:Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a

A:Experimental source: clone ZK829

C:Comment: Lack of unc-22 leads to a constant twitching of the body muscles.

C:Genetics:

A:Gene: unc-22; CESP:ZK617.1a

A:Map position: 4

R:Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 601/3; 691/3; 6776/1; 6808/3

C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

C:Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine/threonine kinase

F:806-898, 899-990, 991-1083, 1084-1175, 1178-1273, 1474-1567, 1770-1864, 2066-2158, 2358-2450, 2451-2510, 5399-5410

96-5790,6263-6356,6386-6478,6541-6635,6649-6742,6745-6838/Region: motif 2

F:1274-1372,1373-1473,1568-1670,1671-1769,1865-1964,1965-2065,2159-2258,2259-2357,2451-2510

23,4215-4313,4314-4415,4416-4516,4612-4710,4711-4811,4908-5009,5010-5109,5110-5210,5399-5410

F:5940-6197/Domain: protein kinase homology <KIN>

F:5948-5956/Region: protein kinase ATP-binding motif

F:5971/Active site: Lys #status predicted

Query Match 19.1%; Score 469; DB 2; Length 6839;

Best Local Similarity 3.5%; Pred.No.2.4;

Matches 220; Conservative 92; Mismatches 131; Indels 5900; Gaps 75;

QY 2 GRKMMVR----- 9

Db 459 QSSAMVVEIQSDVEELMKHDKDAEDYQKEEQSQTLQAEYTKRVARRSKSKSPAP 518

QY 10 ----- 9

Db 519 QAKSTTSSEGRQEAEEVHKRSSVRDPDEESQLDEIPSSGLTIPERRRRELLGQVGE 578

QY 10 -----DVPKMFVLISIFLLVS----- 26

Db 579 SDDEVSEISGLPSFAGGKPRRTDDKPK-----KVSIAFVSTNKSDDDPSTPRRSSI 633

QY 27 -----FINCKVMS----- 34

Db 634 DMRRESVQIELEKTSTPLVPSGASGAPKIVEPVENVVETALTCTCKVSGSPAPTR 693

QY 35 ----- 34

Db 694 WFKGSREVISGGRFKHTDQKEHTVALALLKCRSQDEGPTLTITENVHGTDSADVKLLVT 753

QY 35 ----- 34

Db 754 SDNGLDPRAMLKHRSSQAGFQKDGEGGAGGGGKKPMTAEARRQSLPPGKKVKEWDIP 813

QY 35 -----KALYNRP----- 41

Db 814 LPEKTVQQQVDKICEWKCTYSRPNKIRWYKDKRIFSGGLKYKIVIEKNVCTLIINPE 873

QY 42 ----- 41

Db 874 VDDTGKTCANGVPHTAQLTVLEPPMKYSFLNPLNTQEIYRTKQAVLTCKVTPRAPL 933

QY 42 ----- 41

Db 934 VMYRGSKAIQEGDPRFIEKDAVGRCTLTIKEVEEDDQAEWTARITQDVFSKVQVYVEEP 993

QY 42 -----W-----RGLV 46

Db 994 RHTFVPMKSKQVNESDLATLETVDYDKDAEYVWHDKRIDIDGVKPKVSSNRKRLLI 1053

QY 47 LS----- 48

Db 1054 INGAIEDHGEYKCTTKDDRTMAQLIVDAKNKFIVALKDTVEIKDDVTLTCQTKDKTP 1113

QY 49 -----KIGK-----YKLDOL----- 58

Db 1114 GIWFRNGKQISSMPGKPEFQSRNGHTLKGKIEKMEADVYEDQAGLRGSCNVTVLEA 1173

QY 59 -----KLEILR----- 64

Db 1174 EKRPILNWPKKIBAKAGEPCVVKVPFIKGTRRGDPKQILKNKGPIDEMRKLVVII 1233

QY 65 -----QLETT-----ISTKYNVSKQ- 79

Db 1234 KDDVAEIVFKNPQLADTGKVALELGNAGTALAPPELFFVKDKPKPKGPLETK-NVTAG 1292

QY 80 -----PVK-----NLTMTEF----- 90

Db 1293 LDLYVMGTFDPEGAPVRAIYIEMQEGRSGNWKVGETKGTDFKVKDLKHEGKYPVKAL 1352

QY 91 ----- 90

Db 1353 NECGLSDPLTGESVLAKNPYGVPGKPKMMDAIDVDKDHCTLAWBPPBDDGGAPITGYIIE 1412  
QY 91 ----- 90  
Db 1413 RREKSEKDWQGVQTKPCCCLTDKKVVEDKEYLYRVKAVNKAGPGDPCDHGKPIKMKAK 1472  
QY 91 ---POY----- 93  
Db 1473 KASPEFTGGIKDLRLKVGETIKYDVPISGEPLECLWVVGKPLKAVGRVMSSEGRKH 1532  
QY 94 ----- 93  
Db 1533 IMKIENAVRADSGKFTITLKNSSGCDSTATVTVVGRPTPPKGPLDIADVCADGATLSWN 1592  
QY 94 -----YILAG----- 98  
Db 1593 PPDDGGDPLTGYIVEAQMDNKGKYEIVGKVDPNPTTLKYNGLRNKGNKYFRVKAVNNE 1652  
QY 99 -----PIQNY----- 103  
Db 1653 GESEPLSADQYTOIKDPWDEFGKGRPEITDFDADRIDIAWEPHPKGGAPIEIYEV 1712  
QY 104 ----- 103  
Db 1713 DPDTKEWEKVRVDPDTNASISGLKEGKEYQFRVRAVNAKAGPGQSPSEKQIAKPKFIPA 1772  
QY 104 -----SIT-----YLMF----- 110  
Db 1773 WLKHDNLKSIITKAGATVRWEKIGGEPPEVKWPKGNQQLENGILQITDRKNEHTILC 1832  
QY 111 ----- 110  
Db 1833 IPSAMRSDVGBYRLTVKNSHGADEKANKLTVDLRPSKPNGLPVSVDVPEDNLNLSWKPPD 1892  
QY 111 -----DFYSTQ-----LRKPKAYVY----- 125  
Db 1893 DGGEPPIEYVEKELDTATGRWVPCAKVKOTKAHIDGLKGGTQYQFRVKA VNKESGASDAL 1952  
QY 126 -----SOY----- 128  
Db 1953 STDKDTAKNPYDEPGKTGTPDVVDWDADRVSLEWEPKSGDGPITQYVIEKKGKGRD 2012  
QY 129 -----N 129  
Db 2013 WQECGKVGSDQTNABEILGLKEGEYQFRVKA VNAKAGPGASDPGRKVVAKPRNLKPWIDR 2072  
QY 130 HTAKTITFR-----PP----- 140  
Db 2073 EAMKTITIKVGNVDFVVPVGEPPPKKEWIFNEKPVDDQKIRIESEDYKTRFVLRGATR 2132  
QY 141 ----- 140  
Db 2133 KHAGLYTLTATNASGDKHSVEVIVLGPSSPLGLEVSNVYEDRADLEWKVPEDDGGAP 2192  
QY 141 -----PCGR----- 144  
Db 2193 IDHYEIEKMDLATGRWVPCGRSETTKTVPNLQPGHEYKFRVRAVNAKESGSDPLTTNTAI 2252  
QY 145 -----VPSMTCUSEMLN----- 156  
Db 2253 LAKNPYEPGVKVPDELVDKOHVDLAWNAPDDGGAPFAFVIEKKDKNGRWEALVVP 2312  
QY 157 -----VSKRNDTG----- 164  
Db 2313 GDQKTATVNLKEGEYQFRISARNKAGTDPSPDRVAKPRNLAPRIHREDLSDTTV 2372  
QY 165 -----EQCGNFT----- 172  
Db 2373 KVGATLKPIVHIDGEPAPDVTWSFNGKIGESKAQIENEPYISRFALPKLRKQSGKTYI 2432  
QY 173 -----TFN-----PM-FFNVPR 183  
Db 2433 TATNINGDSVTINI KVKSKTPKGPPIEVTVDFEDRATLDWKPPEDDGGEPPIEYIEK 2492

QY 184 WNTK--LYV-----GPTKVNVD----- 199  
Db 2493 MNTKOGIWWPCGRSGDTHFTVDSLNGDHYKFRKAVNSEGSDPLETETDILAKNPDR 2552  
QY 200 -----OTIYF 204  
Db 2553 PDRGPRPEPTDWDSDHVDLKWDPPLSDGGAPIBBYQIEKRTYGRWEPAITVPGGOT--- 2609  
QY 205 LGLTA----- 209  
Db 2610 ---TATVPDLTPNEEYEFRRVAVNKGSPSDASKAVIAKPRNLKPHIDRDALKNLTIK 2666  
QY 210 -----LLRYAQRNCTHGF- 223  
Db 2667 AGQISFDVPVSGSPAPTIVTWHWPNREIRNGRVRKLDNPEYQSKLVVKQMERGDSGTFT 2726  
QY 224 ----- 223  
Db 2727 IKAVNANGEDATVKINVIDKPTSPNGPLDVSDVHGDHVTLWNRAPDDGGIPIENYVIE 2786  
QY 224 ----- 223  
Db 2787 KYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVA VNAEGESDPLETFTGTTAKDP 2846  
QY 224 ----- 223  
Db 2847 FDKPKTNAPEITDWDKDHVDLEWKPPANDGGAPIEYVVMKDEFPSPMNDVAHVPAQ 2906  
QY 224 ----- 223  
Db 2907 TNATVGNLKEGSKYEFIRAKNKAAGLGDPSASAVAKARNVPPVIDRNSIOEIKVKAGQ 2966  
QY 224 -----YLVNAM 229  
Db 2967 DFLNIPVSGEPTTITWTFEGTVPESDDRMKLNEDGKTKFHVKRALRSDTGYIIKAE 3026  
QY 230 SRN----- 232  
Db 3027 NENGDTAEVKVTVLDHPSSPRGLDVTNIVKDCDLAWKEPEDDGGAEISHYVIEKQDA 3086  
QY 233 ----- 232  
Db 3087 ATGRWTAGESKOTNFHVDLDTQGHYKFRVKA VNRHGDSDPLEAREAIIAKDPDRADK 3146  
QY 233 ----- 232  
Db 3147 PGTPEIVDWDKDHADLKWTPPADGGAPIEGYLVEMRTPSGDWVPVAVTVGAGELTATVDG 3206  
QY 233 -----LFRV-----PKYI----- 240  
Db 3207 LKPGQTYQFRVKA LNKAGESTPSPSRMTWAKPRHLAPKINRDMFVQVRKAGQTLNFDV 3266  
QY 241 -----NGTKL--KOTMR----- 250  
Db 3267 NVEGEPAPKIEWFLNGSLSSGGNTHIDNNNTDNNTKLTTSKTARADSGKYKIVATNESGK 3326  
QY 251 ----- 250  
Db 3327 DEHEVDVNIIDIPCAPSGPLRHKDITKESVVLKWDDBLDDGGSPITVYVVEKQEDGGRWV 3386  
QY 251 -----KLK----- 253  
Db 3387 PCGETSOTSLKVNKLSEGEYKFRVKA VNRQGTSAPLTSDHAI VAKNPFDEPDAPTDVTP 3446  
QY 254 ----- 253  
Db 3447 VMDKDHVDLEWKPPANDGGAPIDAYIVEKKDFGWECARVDGKTTKATADNLTPGET 3506  
QY 254 -----RKOAP----- 258  
Db 3507 YQFRVKA VNAKAGPKSPDPTGNVVAKPRRMAPKUNLAGLLDLRIKAGTPIKLDIAFEGEP 3566

QY 259 ----- 258  
Db 3567 APVAKWANDATIDTGARADVNTPTSSAIHIFSAVRGDTGYKIIIVNEHGKTAQCNV 3626  
QY 259 ----- 258  
Db 3627 TVLDVFTBGPILKIDEIHKEGCTLNWKPTDNGGTDVLHYIVEKMDISRGTWBVGTFP 3686  
QY 259 ----- VKQF----- 263  
Db 3687 DCTAKYNKLVGKEYAFRVAKNVLOESKPLAEBPIIAKNQFVDPDPVKPEVTDMDKD 3746  
QY 264 ----- EKKAKK-----TQSTTTPYFSYTTSAALNVTTWY----- 294  
Db 3747 RIDIKWNPANNGAPVTGYIIVEKKEGSAIWEAGTKP-----GTFPSADNLKPGVEYEF 3803  
QY 295 -----SITTAARV----- 303  
Db 3804 VIAVNAAGSPSDPTDPOITKARYLKPILITASRKIKIRAGFTHNLEVDFIGADPTAT 3863  
QY 304 -----STSTIAY----- 310  
Db 3864 WTVGDSGAALAPELLVDAKSTTSIFPSAKRADSGNYKLVKNELGEDEAIFEIVQDR 3923  
QY 311 ----- 310  
Db 3924 PSAPGPLEVSDVTKDSCVLNWKPPKDDGGAIESNVVVEKRDTKTNTVPSAFVTGTSI 3983  
QY 311 ----- 310  
Db 3984 TVPKLTGHEYEYFRMAENTFGRSDSLNTEPVLAKDPFGTGPKGORPEIVTDNDHIDI 4043  
QY 311 ----- 310  
Db 4044 KWDPPRNGSPVDHYDIERDAKTGRWIKVNTSPVQGTAFSDTRVQKHGTYEYRVAVN 4103  
QY 311 -----RPDSFMK 318  
Db 4104 KAGPGQPSDSSAAATAKPMHEAPKFDLDLGKEFRVKAGEPLVITPFTASQPDISWTK 4163  
QY 319 S----- 319  
Db 4164 EGKPLAGVETSDSQTLVIPSTRSDSGPVKIKAVNPYGEAEANIKITVIDKCAPENI 4223  
QY 320 ----- 319  
Db 4224 TYPVSRHTCTLNWDAPKDDGGAETAGYKIEYQEVGSQIWDKVPGLISGTAYTVRGLHG 4283  
QY 320 ----- 319  
Db 4284 QOYRFRIRABNAVLSDYCGVPVVIKOPFDPGAPSTPEITGYDTNQVSLAWNPPRDDG 4343  
QY 320 -----IMATQ----- 324  
Db 4344 GSPILGVYVERFEKGGDWAPVMPVMVKGTECIVPLGHENETYQFRVAVNAAGHGEPS 4403  
QY 325 ----- 324  
Db 4404 NGSEPVTCRYYVEKPGAPDAPRVGKIKTNSAELTNRLDRDGGAPIDGYIVEKKLGDND 4463  
QY 325 -----LRDLA----- 329  
Db 4464 WTRCNDKPVRTAPEVKNLGEKEYEYFRVAVNSAGEBSPKPSDLVLIEBQPGRPIDFI 4523  
QY 330 ----- 329  
Db 4524 NNLKDI TVRAGETIQIRIPVAGGNPKPIIDLFNNGSPIFENERTVVDVNPGEIVITTGS 4583  
QY 330 -----TW----- 331  
Db 4584 KRSAGPYKISATNYGKDTCKLVNFVLDAPGKPTGPIRATDIOADAMTLSWRPKDNGG 4643  
QY 332 -----VYTTLYR----- 339

Db 4644 DAITNVVSKRTPGGDMVTGHPGVGTTLRVRLNDANTPYEFRVRAENQYGVGEPLETDDA 4703  
QY 340 -----QNPFCBP----- 346  
Db 4704 IVAKNPFDTPGAPGQPEAVEISEAITLQWTRPTSDGGAPIQGYVIEKREVGSTEWTKAA 4763  
QY 347 -----SRNRTA----- 352  
Db 4764 FGNILDTKHRVTGTLTPPKTYEYFRVAAVNAAGQGEYSVNSVPIITADNAPTRPKINMGMLTR 4823  
QY 353 ----- 352  
Db 4824 DILAVAGERAKILVPPFAASPAPKVTFSGENKISPTDPRVKVEYSDFLATLTIEKSELTD 4883  
QY 353 ----- 352  
Db 4884 GGLYFVELENSQSDSDASIRLKVVDPKASPQHIRVEDIAPDCCTLYWMPSPSSGGSPITN 4943  
QY 353 -----VSEFMKNTHVL-----INRTPYTI----- 372  
Db 4944 YIVEKLDLRHSDGKWEKVSSFVRNLNTVGGLIKONRYRFRVRAETQYGVSEPCELADV 5003  
QY 373 -----YG----- 374  
Db 5124 GRNHCTITWMAPLEDGGSKITGVNVEIREYGSTLWTVASDYNVREPEFTVDKLFNFNDE 5183  
QY 375 ----- 374  
Db 5184 FRVAVINAAGKIPSLPSGPIKIOBSSGSRPOIVVKBEDTAQYNRRRAVFTCEAVGREP 5243  
QY 375 -----TLDMSLYYNETMFV----- 389  
Db 5244 TARWLENGRELPESSRYRFEASDGVYKFTIKEVWINDIAGEYTVESVNFYSDTATANLV 5303  
QY 390 ----- 389  
Db 5304 QAPPVIEKDVNTIILPSGDLVRLKIYFSGTAPFRHSLVLRNREIDMDHPTIRIVEFDHI 5363  
QY 390 -----ENKTASDSNKT----- 401  
Db 5364 LITIPALSVREAGRYEYTVSNDGSEATTFWLVNTGLPEAPOGLHISNIGPSTATLSWR 5423  
QY 402 ----- 401  
Db 5424 PVTDDGSKITSYVVEKRDLSKDEWTVTSNVKDMNYIVTGLFENHEYEYFRVSAQNENGI 5483  
QY 402 -----PTSP-----SMGFORTFIDPLWD----- 419  
Db 5484 GAPLVSEHPILARLPDPPTSPNLNLEIVQYGGDYVTLWSOR-----PLSDGGGRLRGYIVE 5539  
QY 420 -----YDLSLLF----- 426  
Db 5540 QOBEHDEWFRCONPSPNNYVNPNLIDGRKYRYRFAVNDAGLSDLAELDQTLFQASG 5599  
QY 427 ----- 426  
Db 5600 SGEPKIVSPLSLNBEVGRCVTFBCEISGSPREYRWFPGCKELVDTSKYTLINKDKQ 5659  
QY 427 ----- 426  
Db 5660 VLINDLTDDADEYTCRATNSGSTRANLRITKPRVFPFPKYVGYEAKGETIEL 5719  
QY 427 ----- 426



Db 5720 KIPYKAYPQGEARWTKDGEKIENNSKFSITTTDDKFATLIRISNASREDYGEYRVVVSVSG 5779  
Qy 427 -----LDEI----- 430  
|||  
Db 5780 SDSGTVNTVADVPPPPPIIENILDEAVILSWKPPALDGGSLVNTVIEKREAMGGSW 5839  
Qy 431 ----- 430  
Db 5840 SPCAKSRYTYTIEGLRAGKQYEFRIIAENKHGSKPCPTAPVLIPGDERKRRGYDVD 5899  
Qy 431 ----- 430  
Db 5900 EGGKIVRGKGTSSNDYVFDIWKQYQPQVBIKHDHVLHDYDIHEELGTGAGGVHRV 5959  
Qy 431 -----RNFS-----LRSPTYVNL----- 443  
|||  
Db 5960 TERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAPEDDNEMWMIYEFM 6019  
Qy 444 ----- 443  
Db 6020 SGGELFEKVADEHNKMSDEAVEYMRQVCKGLCHHENNYVHLDLKPENIMFTTKRSNEL 6079  
Qy 444 ----- 443  
Db 6080 KLIDFGLTAHLDPKQSVKVTGTAEFAAPEVAEGKPGVGYTDMMSVGVLSYLLSGLSPF 6139  
Qy 444 ----- 443  
Db 6140 GGENDETIRNVKSCDWNDDSAFSGISEDGKDFIRKILLADPNTRMTIHQALEHPWLTP 6199  
Qy 444 -----Tpp-----EHR----- 449  
|||  
Db 6200 GNAFGRDSQIPSSRYTKIRDSIKTKYDAWPEPLPLGRISNYSSLRKHQPQEYSIRDAPW 6259  
Qy 450 ----- 449  
Db 6260 DRSEAQRPIVVKPYGTEVGEQSANFYCRVYASSPPVVVTHKDDRELKQSVKMKRYNGN 6319  
Qy 450 -----RAVN----- 453  
|||  
Db 6320 DYGLTINRVKDDKGEYTVRAKNISYGTKEIVFLNVTNRHSPLKPEPLEPMKAPSPRV 6379  
Qy 454 ----- 453  
Db 6380 EEFKERRSAPPTFHLNRLLIQKHQCKLTCLOGNPNTIEMWKDGHVPDEDRVQVSFR 6439  
Qy 454 ----- 453  
Db 6440 SGVCSLEIFNARVDDAGTYTATNDLGVDSBVCVLTVQTKGGEPIPRVSSFRPRRAYDT 6499  
Qy 454 LST----- 456  
|||  
Db 6500 LSTGTDVERSHYADMRRLIRDVSPDVSRAADDLTKITNELPSPFTAQLSDSETEVGG 6559  
Qy 457 ----- 456  
Db 6560 SAEPFAAVSGQPEPLIEWLHNGERISEDSRFRASYVAGKATLRISDAKSDGQYL CRA 6619  
Qy 457 SNS----- 459  
|||  
Db 6620 SNSAQEQTRATLVKGDQPLNGHAGQAVSELVRVKHLGGEIVNNGESVTFEAVQGT 6679  
Qy 460 -----LW----- 461  
|||  
Db 6680 PEEVLWNRNGELTNGDKTISQDGETLSFTINSADASDAGHYQLEVRKGNLVSASL 6739  
Qy 462 -----WV 463  
|||  
Db 6740 VVVGKADPPVTRLPSVSAPIGGSTAFTIEFENVEGLTVQWP 6782



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:22:26 ; Search time 196 Seconds  
(without alignments)  
1365.047 Million cell updates/sec

Title: US-09-942-146a-1  
Perfect score: 2451  
Sequence: 1 MGRKEMVRDVPKMFVLISI.....PEHRRVNLSTNSLWMLQ 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.0

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2417	98.6	466	1 UL74 HCMVA	P16750 human cytom
2	2417	98.6	466	2 Q7M618	Q7m618 human cytom
3	2417	98.6	466	2 AAN40055	Aan40055 human cytom
4	2411	98.4	466	2 Q8AZ28	Q8az28 human cytom
5	2407	98.2	466	2 Q8BCU8	Q8bcu8 human cytom
6	2398	97.8	466	2 Q8AZ36	Q8az36 human cytom
7	2398	97.8	466	2 Q8BCU2	Q8bcu2 human cytom
8	2071	84.5	463	2 Q8AZ32	Q8az32 human cytom
9	2069	84.4	464	2 Q8BCU3	Q8bcu3 human cytom
10	2068	84.4	463	2 Q8BCU9	Q8bcu9 human cytom
11	2068	84.4	463	2 Q8BCV2	Q8bcv2 human cytom
12	2067	84.3	463	2 Q8AZ39	Q8az39 human cytom
13	2057	83.9	464	2 Q8BCU8	Q8bcu8 human cytom
14	2057	83.9	464	2 AAS48965	Aas48965 human cytom
15	2054	83.8	462	2 Q8BCU0	Q8bcu0 human cytom
16	2049	83.6	462	2 Q8BCU6	Q8bcu6 human cytom
17	1947	79.4	465	2 Q8AYW0	Q8ayw0 human cytom
18	1941	79.2	465	2 Q8BCU9	Q8bcu9 human cytom
19	1906	77.8	462	2 Q8AYZ3	Q8ayz3 human cytom
20	1901	77.6	462	2 Q8BCU5	Q8bcu5 human cytom
21	1877	76.6	472	2 Q8BCV1	Q8bcv1 human cytom
22	1877	76.6	472	2 AAR31626	Aar31626 human cytom
23	1759	71.8	464	2 Q8BCU1	Q8bcu1 human cytom
24	1751	71.4	464	2 Q8BCU4	Q8bcu4 human cytom
25	1748	71.3	464	2 Q8AYV5	Q8ayv5 human cytom
26	1746	71.2	464	2 Q8AYZ5	Q8ayz5 human cytom
27	1740	71.0	457	2 Q8BCU7	Q8bcu7 human cytom
28	1740	71.0	457	2 Q8BCV0	Q8bcv0 human cytom
29	1026	41.9	461	2 Q8QS25	Q8qs25 chimpanzee
30	820	33.5	389	2 Q7TFN1	Q7tfn1 thesus cyto
31	718	29.3	144	2 Q7T9B3	Q7t9b3 human cytom

#### ALIGNMENTS

##### RESULT 1

ID	UL74 HCMVA	STANDARD;	PRT;	466 AA.
AC	P16750;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DB	Glycoprotein UL74 precursor.			
GN	Name=UL74;			
OS	Human cytomegalovirus (strain AD169).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Betaherpesvirinae; Cytomegalovirus.			
OX	NCBI_TaxID=10360;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90269039; Pubmed=2161319;			
RA	Chee N.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,			
RA	Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,			
RA	Predlie B., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;			
RT	"Analysis of the protein-coding content of the sequence of human			
RT	cytomegalovirus strain AD169."			
RL	Curr. Top. Microbiol. Immunol. 154:125-169(1990).			
CC	-1- SIMILARITY: Belongs to the herpesviruses U47 family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	X17403; CAA35389.1; -			
DR	PIR; S09837; S09837.			
KW	Glycoprotein; Signal.			
FT	SIGNAL 1 30			
FT	CHAIN 31 466			Potential.
FT	CARBOHYD 75 75			Glycoprotein UL74.
FT	CARBOHYD 83 83			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 87 87			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 103 103			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 130 130			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 157 157			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 162 162			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 171 171			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 219 219			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 242 242			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 288 288			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 292 292			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 350 350			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 385 385			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 392 392			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 399 399			N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD 433 433			N-linked (GlcNAc. .) (Potential).

Q7t9b4 human cytom  
Q7t9b5 human cytom  
Q7t9b6 human cytom  
Q7t9c3 human cytom  
Q7t9c4 human cytom  
Q7t9c5 human cytom  
Q7t9d2 human cytom  
Q7t9d4 human cytom  
Q7t9d5 human cytom  
Q8wz42 homo sapien  
Q7t9b2 human cytom  
Q7t9c1 human cytom  
Q10466 homo sapien  
Q8njx1 trichoderma

```

FT CARBOHYD 454 454 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

Query Match 98.6%; Score 2417; DB 1; Length 466;
Best Local Similarity 98.9%; Pred. No. 8.8e-58;
Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVMDVPMFVLSISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
DB 1 MGRKEMVMDVPMFVLSISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
QY 61 EILRQLETTISTKYNVSKQPVKNLTMN-TEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQPVKNLTMNTEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 120
QY 120 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 179
DB 121 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 180
QY 180 NVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLNVNMSRNLFRVPKY 239
DB 181 NVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLNVNMSRNLFRVPKY 240
QY 240 INGTKLNTWRKLRKQAPVKEQFEKAKKTOSTTTPYPSYTTSAALNVTTNVTYSITTA 299
DB 241 INGTKLNTWRKLRKQAPVKEQFEKAKKTOSTTTPYPSYTTSAALNVTTNVTYSITTA 300
QY 300 ARRVSTSTIAYRPDSSFMKSIMATQRLDIAWTWYTLRYRQNPFCPSRNRNAVSEFMKN 359
DB 301 ARRVSTSTIAYRPDSSFMKSIMATQRLDIAWTWYTLRYRQNPFCPSRNRNAVSEFMKN 360
QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 419
DB 361 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSLSRSTYVNLTPPEHRRVAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSLSRSTYVNLTPPEHRRVAVNLSTNSLWMLQ 466

RESULT 2
Q7M6L8 PRELIMINARY; PRT; 466 AA.
ID Q7M6L8
AC Q7M6L8;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ad169;
RX MEDLINE=22255414; PubMed=12533697;
RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
RA "The genes encoding the gCIII complex of human cytomegalovirus exist
RA in highly diverse combinations in clinical isolates.";
RA Alcendor D.J., McGeoch D.J., Hayward G.S.;
RT "The human cytomegalovirus genome revisited: comparison with the
RT chimpanzee cytomegalovirus genome.";
RL J. Gen. Virol. 84:17-28(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ad169;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
RA "The genes encoding the gCIII complex of human cytomegalovirus exist
RA in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; BK000394; DAA00170.1; -
DR EMBL; AF531331; AAN40055.1; -
SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

```

```

Query Match 98.6%; Score 2417; DB 2; Length 466;
Best Local Similarity 98.9%; Pred. No. 8.8e-58;
Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVMDVPMFVLSISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
DB 1 MGRKEMVMDVPMFVLSISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
QY 61 EILRQLETTISTKYNVSKQPVKNLTMN-TEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQPVKNLTMNTEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 120
QY 120 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 179
DB 121 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 180
QY 180 NVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLNVNMSRNLFRVPKY 239
DB 181 NVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLNVNMSRNLFRVPKY 240
QY 240 INGTKLNTWRKLRKQAPVKEQFEKAKKTOSTTTPYPSYTTSAALNVTTNVTYSITTA 299
DB 241 INGTKLNTWRKLRKQAPVKEQFEKAKKTOSTTTPYPSYTTSAALNVTTNVTYSITTA 300
QY 300 ARRVSTSTIAYRPDSSFMKSIMATQRLDIAWTWYTLRYRQNPFCPSRNRNAVSEFMKN 359
DB 301 ARRVSTSTIAYRPDSSFMKSIMATQRLDIAWTWYTLRYRQNPFCPSRNRNAVSEFMKN 360
QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 419
DB 361 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSLSRSTYVNLTPPEHRRVAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSLSRSTYVNLTPPEHRRVAVNLSTNSLWMLQ 466

RESULT 3
AAN40055 PRELIMINARY; PRT; 466 AA.
ID AAN40055
AC AAN40055;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ad169;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
RA "The genes encoding the gCIII complex of Human Cytomegalovirus Exist
RA in Highly Diverse Combinations in Clinical Isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531331; AAN40055.1; -
SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

Query Match 98.6%; Score 2417; DB 2; Length 466;
Best Local Similarity 98.9%; Pred. No. 8.8e-58;
Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVMDVPMFVLSISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
DB 1 MGRKEMVMDVPMFVLSISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
QY 61 EILRQLETTISTKYNVSKQPVKNLTMN-TEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQPVKNLTMNTEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 120
QY 120 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 179

```

```

Db 121 PAKVYSQYNHTAKITIFRPPCGTVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVRWNTKLYVGPVKVNDSTQIIFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
Db 181 NVRWNTKLYVGPVKVNDSTQIIFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
Qy 240 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQRLD L ATWVYTLRYQNPFCEPSRNR TAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQRLD L ATWVYTLRYQNPFCEPSRNR TAVSEFMKN 360
Qy 360 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFOR TIDPLWD 419
Db 361 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFOR TIDPLWD 420
Qy 420 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 466

```

## RESULT 4

```

Q8AZ28 ID Q8AZ28 PRELIMINARY; PRT; 466 AA.
AC Q8AZ28;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 05-JUN-2004 (Tremblrel. 27, Last annotation update)
DE UL74 protein.
OS Human Cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=298, 6444, and SW3;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531320; AAN40044.1; -
DR EMBL; AF531325; AAN40049.1; -
DR EMBL; AF531337; AAN40061.1; -
DR EMBL; AF531346; AAN40070.1; -
SQ SEQUENCE 466 AA; 54200 MW; 56581B75919572FE CRC64;

```

```

Query Match 98.4%; Score 2411; DB 2; Length 466;
Best Local Similarity 98.7%; Pred. No. 1.3e-57;
Matches 460; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MGRKEMVVRDVPKMFVLISIFLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
Db 1 MGRKEMVVRDVPKMFVLISIFLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
Qy 61 EILRQLETTISTKYNVSKQPVKNTLMN-TEPPQYIILAGPIQNYSIYLVWDFDYSTOLRK 119
Db 61 EILRQLETTISTKYNVSKQPVKNTLMNTEPPQYIILAGPIQNYSIYLVWDFDYSTOLRK 120
Qy 120 PAKVYSQYNHTAKITIFRPPCGRVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYSQYNHTAKITIFRPPCGTVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVRWNTKLYVGPVKVNDSTQIIFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
Db 181 NVRWNTKLYVGPVKVNDSTQIIFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
Qy 240 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 EILRQLETTISTKYNVSKQPVKNTLMNTEPPQYIILAGPIQNYSIYLVWDFDYSTOLRK 120
Qy 120 PAKVYSQYNHTAKITIFRPPCGRVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYSQYNHTAKITIFRPPCGTVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVRWNTKLYVGPVKVNDSTQIIFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
Db 181 NVRWNTKLYVGPVKVNDSTQIIFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
Qy 240 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300

```

```

Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQRLD L ATWVYTLRYQNPFCEPSRNR TAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQRLD L ATWVYTLRYQNPFCEPSRNR TAVSEFMKN 360
Qy 360 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFOR TIDPLWD 419
Db 361 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFOR TIDPLWD 420
Qy 420 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 466

```

## RESULT 5

```

Q8BCU8 ID Q8BCU8 PRELIMINARY; PRT; 466 AA.
AC Q8BCU8;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE UL74 protein.
OS Human Cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=851;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531330; AAN40054.1; -
SQ SEQUENCE 466 AA; 54230 MW; 56581B75919562FF CRC64;

```

```

Query Match 98.2%; Score 2407; DB 2; Length 466;
Best Local Similarity 98.5%; Pred. No. 1.7e-57;
Matches 459; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 MGRKEMVVRDVPKMFVLISIFLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
Db 1 MGRKEMVVRDVPKMFVLISIFLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
Qy 61 EILRQLETTISTKYNVSKQPVKNTLMN-TEPPQYIILAGPIQNYSIYLVWDFDYSTOLRK 119
Db 61 EILRQLETTISTKYNVSKQPVKNTLMNTEPPQYIILAGPIQNYSIYLVWDFDYSTOLRK 120
Qy 120 PAKVYSQYNHTAKITIFRPPCGRVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYSQYNHTAKITIFRPPCGTVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVRWNTKLYVGPVKVNDSTQIIFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
Db 181 NVRWNTKLYVGPVKVNDSTQIIFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
Qy 240 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQRLD L ATWVYTLRYQNPFCEPSRNR TAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQRLD L ATWVYTLRYQNPFCEPSRNR TAVSEFMKN 360
Qy 360 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFOR TIDPLWD 419
Db 361 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFOR TIDPLWD 420
Qy 420 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 466

```

```
RESULT 6
Q8AZ36
ID Q8AZ36 PRELIMINARY; PRT; 466 AA.
AC Q8AZ36;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4088, and SW1762;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531322; AAN40046.1; -.
DR EMBL; AF531344; AAN40068.1; -.
SQ SEQUENCE 466 AA; 54215 MW; 4DC855844E3E65A0 CRC64;

Query Match 97.8%; Score 2398; DB 2; Length 466;
Best Local Similarity 98.3%; Pred. No. 3e-57;
Matches 458; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MGRKEMVRDVPKMFVLISFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYLQDLKL 60
DB 1 MGRKEMVRDVPKMFVLISFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYLQDLKL 60
QY 61 EILRQLETTISTKYNVSKQVPKVLWN-TFFPOYIILAGPIQNYSTIYLFDFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQVPKVLWN-TFFPOYIILAGPIQNYSTIYLFDFYSTQLRK 120
QY 120 PAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 PAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
QY 180 NVPRWNTKLYVGTQKVNDSQIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 239
DB 181 NVPRWNTKLYVGTQKVNDSQIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 240
QY 240 INGTKLNTMRKLKRQAPVKEQFEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
DB 241 INGTKLNTMRKLKRQAPVKEQFEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
QY 300 ARRVSTTIAYRPDSFMSKINATQRLDRLATWVYTLRYRONPCEPSRNRRTAVSEFMKN 359
DB 301 ARRVSTTIAYRPDSFMSKINATQRLDRLATWVYTLRYRONPCEPSRNRRTAVSEFMKN 360
QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 419
DB 361 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSRLSPTVYVNLTPPEHRAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSRLSPTVYVNLTPPEHRAVNLSTNSLWMLQ 466

RESULT 7
Q8BCU2
ID Q8BCU2 PRELIMINARY; PRT; 466 AA.
AC Q8BCU2;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4088, and SW1762;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531322; AAN40046.1; -.
DR EMBL; AF531344; AAN40068.1; -.
SQ SEQUENCE 466 AA; 54215 MW; 4DC855844E3E65A0 CRC64;

Query Match 97.8%; Score 2398; DB 2; Length 466;
Best Local Similarity 98.3%; Pred. No. 3e-57;
Matches 458; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MGRKEMVRDVPKMFVLISFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYLQDLKL 60
DB 1 MGRKEMVRDVPKMFVLISFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYLQDLKL 60
QY 61 EILRQLETTISTKYNVSKQVPKVLWN-TFFPOYIILAGPIQNYSTIYLFDFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQVPKVLWN-TFFPOYIILAGPIQNYSTIYLFDFYSTQLRK 120
QY 120 PAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 PAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
QY 180 NVPRWNTKLYVGTQKVNDSQIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 239
DB 181 NVPRWNTKLYVGTQKVNDSQIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 240
QY 240 INGTKLNTMRKLKRQAPVKEQFEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
DB 241 INGTKLNTMRKLKRQAPVKEQFEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
QY 300 ARRVSTTIAYRPDSFMSKINATQRLDRLATWVYTLRYRONPCEPSRNRRTAVSEFMKN 359
DB 301 ARRVSTTIAYRPDSFMSKINATQRLDRLATWVYTLRYRONPCEPSRNRRTAVSEFMKN 360
QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 419
DB 361 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSRLSPTVYVNLTPPEHRAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSRLSPTVYVNLTPPEHRAVNLSTNSLWMLQ 466

RESULT 8
Q8AZ32
ID Q8AZ32 PRELIMINARY; PRT; 463 AA.
AC Q8AZ32;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=279, and SW2;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531319; AAN40043.1; -.
DR EMBL; AF531345; AAN40069.1; -.
SQ SEQUENCE 463 AA; 53871 MW; 51310C0D1D239A5A CRC64;
```

Query Match 84.5%; Score 2071; DB 2; Length 463;  
 Best Local Similarity 85.4%; Pred. No. 3.4e-48;  
 Matches 399; Conservative 22; Mismatches 40; Indels 6; Gaps 4;

QY 1 MGRKEMVDRVPMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLK 60  
 DB 1 MGRKEM-RGVNLFLLSFLVSPINCRVAVLSVGRVWSGVLSIGKQRLDKFL 59

QY 61 EILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYISITLWFDYFYSTQLR 118  
 DB 60 EILQLEKDIYTKYFNMTROHINKLTMNTEPPRYIILAGPIQNNSTVYLWFDYFYSTQLR 119

QY 119 KPAKYVYQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMF 178  
 DB 120 KPAKYVSEYNHTAKTITFRPPSCGTVPSPMTCLSEMLNVSKRNDTGEQCGNFTTFNPMF 179

QY 179 FNVPRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 238  
 DB 180 FNVPRWNTKLYGSKKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 239

QY 239 YINGTKLNTMRKLRKQAPVKEQFEKAKKTQSTTTTPYFYSYTTSAALNVTNTVYITT 298  
 DB 240 YINGTKLNTMRKLRKQAPVKEQSEKSKSQSTTTTPYSPYTTSTALNVTNTVYITT 299

QY 299 AARRVSTSTIAYRDPSSFMKSIMATQRLDRLATWYVYTLRYQNPCEPSRNRRTAVSEPMK 358  
 DB 300 TARRVSTSTIAYRDPSSFMKSIMTQRLDRLATWYVYTLRYQNPCEPSRNRRTAVSEPMK 359

QY 359 NTHVLIRNETPTIYIGTLDMSLYNETMVENKTASDNKTTPSPSMGFQRTFIDPLW 418  
 DB 360 NTHVLIRNETPTIYIGTLDMSLYNETMVENKTASB---TTTSPSTGTGQRTFIDPLW 416

QY 419 DYLDLSLLFLDEIRNFSRLSPTYVNLTPPEHRAVNLSTNSLWMLQ 465  
 DB 417 DYLDLSLLFLDEIRNFSRLSPTYVNLTPPEHRAVNLSTNSLWMLQ 463

RESULT 9  
 Q8BCU3 ID Q8BCU3 PRELIMINARY; PRT; 464 AA.  
 AC Q8BCU3  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE UL74 protein.  
 OS Human cytomegalovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OX NCBI\_TaxID=10359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SW1324;  
 RX MEDLINE=2255414; PubMed=12368327;  
 RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;  
 RT "The genes encoding the gCIII complex of human cytomegalovirus exist  
 in highly diverse combinations in clinical isolates."  
 RL J. Virol. 76:10841-10848(2002).  
 DR EMBL; AF531340; AAN40064.1; -.  
 SQ SEQUENCE 464 AA; 54146 MW; A24A513BF112848B CRC64;

Query Match 84.4%; Score 2069; DB 2; Length 464;  
 Best Local Similarity 84.3%; Pred. No. 3.9e-48;  
 Matches 397; Conservative 25; Mismatches 36; Indels 13; Gaps 5;

QY 1 MGRKEMVDRVPMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLD 56  
 DB 1 MGRKEM-RSISKUFFIISLTVLFSINCKV-----RPPGRVWLGTVLSTIGKQKLD 53

QY 57 QLKLEILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYISITLWFDYFYS 114  
 DB 54 KFKLEILQLEREPTTKYFNMTROHINKLTMNTEPPRYIILAGPIQNDSTIYLWFDYFYS 113

QY 115 TQLRKPAKYVYQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTF 174

DB 114 TQLRKPAKYVYQYNHTAKTITFRPPSCGTVPSPMTCLSEMLNVSKRNDTGEQCGNFTTF 173  
 QY 175 NPMFNVPRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 234  
 DB 174 NPMFNVPRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 233

QY 235 RVPKYINGTKLNTMRKLRKQAPVKEQFEKAKKTQSTTTTPYFYSYTTSAALNVTNTVYITT 294  
 DB 234 RVPKYINGTKLNTMRKLRKQAPVKEQLEKTKKSQSTTTTPYFYSYTTSTALNVTNTVYITT 293

QY 295 SITTAARRVSTSTIAYRDPSSFMKSIMATQRLDRLATWYVYTLRYQNPCEPSRNRRTAVS 354  
 DB 294 RVTTSARRVSTSTIAYRDPSSFMKSIMATQRLDRLATWYVYTLRYQNPCEPSRNRRTAVS 353

QY 355 EFMKNTHVLIRNETPTIYIGTLDMSLYNETMVENKTASDNKTTPSPSMGFQRTFIDPLW 414  
 DB 354 EFMKNTHVLIRNETPTIYIGTLDMSLYNETMVENKTASDNKTTPSPSTRQKTFI 413

QY 415 DPLWDYLDLSLLFLDEIRNFSRLSPTYVNLTPPEHRAVNLSTNSLWMLQ 465  
 DB 414 DPLWDYLDLSLLFLDKIRNFSRLSPAYGNLTPEHRAVNLSTNSLWMLQ 464

RESULT 10  
 Q8BCU3 ID Q8BCU3 PRELIMINARY; PRT; 463 AA.  
 AC Q8BCU3  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE UL74 protein.  
 OS Human cytomegalovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OX NCBI\_TaxID=10359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SW990;  
 RX MEDLINE=2255414; PubMed=12368327;  
 RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;  
 RT "The genes encoding the gCIII complex of human cytomegalovirus exist  
 in highly diverse combinations in clinical isolates."  
 RL J. Virol. 76:10841-10848(2002).  
 DR EMBL; AF531354; AAN40078.1; -.  
 SQ SEQUENCE 463 AA; 53885 MW; E86783A05FB9424A CRC64;

Query Match 84.4%; Score 2068; DB 2; Length 463;  
 Best Local Similarity 85.2%; Pred. No. 4.1e-48;  
 Matches 398; Conservative 23; Mismatches 40; Indels 6; Gaps 4;

QY 1 MGRKEMVDRVPMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGKYKLDQLK 60  
 DB 1 MGRKEM-RGVNLFLLSFLVSPINCRVAVLSVGRVWSGVLSIGKQRLDKFL 59

QY 61 EILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYISITLWFDYFYSTQLR 118  
 DB 60 EILQLEKDIYTKYFNMTROHINKLTMNTEPPRYIILAGPIQNNSTVYLWFDYFYSTQLR 119

QY 119 KPAKYVYQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMF 178  
 DB 120 KPAKYVSEYNHTAKTITFRPPSCGTVPSPMTCLSEMLNVSKRNDTGEQCGNFTTFNPMF 179

QY 179 FNVPRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 238  
 DB 180 FNVPRWNTKLYGSKKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 239

QY 239 YINGTKLNTMRKLRKQAPVKEQFEKAKKTQSTTTTPYFYSYTTSAALNVTNTVYITT 298  
 DB 240 YINGTKLNTMRKLRKQAPVKEQSEKSKSQSTTTTPYSPYTTSTALNVTNTVYITT 299

QY 299 AARRVSTSTIAYRDPSSFMKSIMATQRLDRLATWYVYTLRYQNPCEPSRNRRTAVSEPMK 358

```

Db 300 TARRVSTSTIAYRPDSSFMKSIWTTQLRDLATWYVTLRYRQNPFCSSRNRTAVSEFMK 359
Qy 359 NTHVLIRNETPYTYIGTLDMSLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418
Db 360 NTHVLIRNETPYTYIGTLDMSLYNETMFVENKTASE--TTPTSPSTGFGQRTFIDPLW 416
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 463

RESULT 11
Q8BCV2
ID Q8BCV2 PRELIMINARY; PRT; 463 AA.
AC Q8BCV2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=122;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531326; AAN40050.1; -.
DR EMBL; AF531334; AAN40058.1; -.
SQ SEQUENCE 463 AA; 53957 MW; 55A7D4157FD32424 CRC64;

Query Match 84.4%; Score 2068; DB 2; Length 463;
Best Local Similarity 85.4%; Pred. No. 4.1e-48;
Matches 399; Conservative 22; Mismatches 40; Indels 6; Gaps 4;

Qy 1 MGRKEMVRDVPKMFVLISLFLVGFINKVMSKALYNRPWRGLVLSKIGYKLDQLK 60
Db 1 MGRKEM- RGVNLFLLMSLTLFLFSPINCAAVLSVGRVMSGKVLSTIGKRLDKPKL 59
Qy 61 EILRQLETTISTKY-NVSKQPVKNLTWN-TEFPQYVILAGPIQNYSTVLMFDFYSTOLR 118
Db 60 EILKOLEKDIYTKYFNNTQHIKNLTWNTEFPRIYVILAGPIQNNSTVLMFDFYSTOLR 119
Qy 119 KPAKYVYVSOVNHAKTITFRPPCGVRPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMF 178
Db 120 KPAKYVSEVNHAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMF 179
Qy 179 FNVPRNWKLYVGPVKVNDVSDTIYFLGLTALLLYAQRNCTHSPYLVNMSRNLFRVPK 238
Db 180 FNVPRNWKLYVGSKKVNDVSDTIYFLGLTALLLYAQRNCTHSPYLVNMSRNLFRVPK 239
Qy 239 YINGTKLKNMVKLRKQAPVKEQEKAKTKQSTTTPYFSTYTTSAALNVTNVTYSITT 298
Db 240 YINGTKLKNMVKLRKQAPVKEQEKSKSKSQSTTTPYSPYTTSTALNVTNVTYSITT 299
Qy 299 AARRVSTSTIAYRPDSSFMKSIWTTQLRDLATWYVTLRYRQNPFCPSRNRTAVSEFMK 358
Db 300 TARRVSTSTIAYRPDSSFMKSIWTTQLRDLATWYVTLRYRQNPFCSSRNRTAVSEFMK 359
Qy 359 NTHVLIRNETPYTYIGTLDMSLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418
Db 360 NTHVLIRNETPYTYIGTLDMSLYNETMFVENKTASE--TTPTSPSTGFGQRTFIDPLW 416
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 463

```

RESULT 12  
Q8AZ39

```

ID Q8AZ39 PRELIMINARY; PRT; 463 AA.
AC Q8AZ39;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=650, and DM7;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531326; AAN40050.1; -.
DR EMBL; AF531334; AAN40058.1; -.
SQ SEQUENCE 463 AA; 53901 MW; 403100AADBF4412A CRC64;

Query Match 84.3%; Score 2067; DB 2; Length 463;
Best Local Similarity 85.2%; Pred. No. 4.4e-48;
Matches 398; Conservative 22; Mismatches 41; Indels 6; Gaps 4;

Qy 1 MGRKEMVRDVPKMFVLISLFLVGFINKVMSKALYNRPWRGLVLSKIGYKLDQLK 60
Db 1 MGRKEM- RGVNLFLLMSLTLFLFSPINCAAVLSVGRVMSGKVLSTIGKRLDKPKL 59
Qy 61 EILRQLETTISTKY-NVSKQPVKNLTWN-TEFPQYVILAGPIQNYSTVLMFDFYSTOLR 118
Db 60 EILKOLEKDIYTKYFNNTQHIKNLTWNTEFPRIYVILAGPIQNNSTVLMFDFYSTOLR 119
Qy 119 KPAKYVYVSOVNHAKTITFRPPCGVRPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMF 178
Db 120 KPAKYVSEVNHAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMF 179
Qy 179 FNVPRNWKLYVGPVKVNDVSDTIYFLGLTALLLYAQRNCTHSPYLVNMSRNLFRVPK 238
Db 180 FNVPRNWKLYVGSKKVNDVSDTIYFLGLTALLLYAQRNCTHSPYLVNMSRNLFRVPK 239
Qy 239 YINGTKLKNMVKLRKQAPVKEQEKAKTKQSTTTPYFSTYTTSAALNVTNVTYSITT 298
Db 240 YINGTKLKNMVKLRKQAPVKEQEKSKSKSQSTTTPYSPYTTSTALNVTNVTYSITT 299
Qy 299 AARRVSTSTIAYRPDSSFMKSIWTTQLRDLATWYVTLRYRQNPFCPSRNRTAVSEFMK 358
Db 300 TARRVSTSTIAYRPDSSFMKSIWTTQLRDLATWYVTLRYRQNPFCSSRNRTAVSEFMK 359
Qy 359 NTHVLIRNETPYTYIGTLDMSLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418
Db 360 NTHVLIRNETPYTYIGTLDMSLYNETMFVENKTASE--TTPTSPSTGFGQRTFIDPLW 416
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 463

RESULT 13
Q8BCT8
ID Q8BCT8 PRELIMINARY; PRT; 464 AA.
AC Q8BCT8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE UL74 protein.
GN Name=UL74; ORFNames=HHV5g069;
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]

```



[illegible][illegible]

Db 60 EILKQLEKDIYKYFNMTQHIKMLTNMTPEPRYYILAGPIQNNSVTYLMWDFYSTQLR 119  
QY 119 KPAKYVYSQVNHAKTITFRPPCGRYPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMF 178  
Db 120 KPAKYVSEYNHAKTITFRPPCGTVPSCGLSEMLNVSKRNDTGEQCGN-TTFNPMF 178  
QY 179 FNVPRWNTKLYVGTQKVNDSQTYIFLGLTALLRYAQRNCTHSFYLNVMSRNLPRYPK 238  
Db 179 FNVPRWNTKLYVGSKKVNDSQTYIFLGLTALLRYAQRNCTHSFYLNVMSRNLPRYPK 238  
QY 239 YINGTKLKNTRKLRKQAPVKEQPEKAKTQSTTTPYFSYTTSAALNVTTNVYSITT 298  
Db 239 YINGTKLKNTRKLRKQAPVKEQSEKSKSQSTTTPYSPYTTSTALNVTTNATYSVTT 298  
QY 299 AARRVSTSTIAYRPDSSFMSKIMATQRLDRLATWVTTILRYRQNPCEPSRNRNAVSEFMK 359  
Db 299 TARRISTSTIAYRPDSSFMSKIMTTLQRLDRLATWVTTILRYRQNPCESSRNRNAVSEFMK 358  
QY 359 NTHVLIRNETPYTYIGTLDMSLYYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLW 418  
Db 359 NTHVLIRNETPYTYIGTLDMSLYYNETMFVENKTASE--TTPTSPSTGFQRTFIDPLW 415  
QY 419 DYLDLSLLFLDEIRNFSLRSPYVNLTPPEHRAVNLSNSTSLWWWLQ 465  
Db 416 DYLDLSLLFLDEIRNFSLSQSPYGNLTPEHRAVNLSNSTSLWWWLQ 462

Search completed: December 7, 2004, 15:39:25  
Job time : 200 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:30:25 ; Search time 40 Seconds  
(without alignments)  
770.947 Million cell updates/sec

Title: US-09-942-146A-1  
Perfect score: 2451  
Sequence: 1 MGRKEMVDRVPMFVLSI.....PEHRAVNLSTNSLWMLQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.0

Searched: 478139 seqs, 56318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pdp.\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pdp.\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pdp.\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pdp.\*  
5: /cgn2\_6/prodata/1/iaa/PCUTUS COMB.pdp.\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2420	98.7	465	US-09-942-986-1	Sequence 1, Appli
2	575	23.5	15281	US-08-471-119A-2	Sequence 2, Appli
3	516	21.1	10182	US-09-134-001C-3159	Sequence 3159, Ap
4	483	19.7	11877	US-09-105-537-6	Sequence 6, Appli
5	449	18.3	4563	US-09-108-006C-1	Sequence 1, Appli
6	449	18.3	4563	US-09-538-092-842	Sequence 842, App
7	440	18.0	4536	US-09-180-422B-27	Sequence 27, Appli
8	440	18.0	4536	US-09-079-030-1	Sequence 1, Appli
9	426	17.4	3898	US-08-750-717-2	Sequence 2, Appli
10	422	17.2	3778	US-08-222-617A-2	Sequence 2, Appli
11	420	17.1	3666	US-08-222-617A-12	Sequence 12, Appli
12	420	17.1	3727	US-08-222-617A-27	Sequence 27, Appli
13	420	17.1	3898	US-08-876-991-2	Sequence 2, Appli
14	420	17.1	3898	US-09-059-853-2	Sequence 2, Appli
15	417	17.0	4654	US-08-476-515A-84	Sequence 84, Appli
16	416	17.0	4655	US-08-652-877-84	Sequence 84, Appli
17	416	17.0	4655	US-08-652-877-86	Sequence 86, Appli
18	416	17.0	4655	US-08-652-877-88	Sequence 88, Appli
19	416	17.0	4655	US-08-652-877-90	Sequence 90, Appli
20	412	16.8	3623	US-09-341-461-2	Sequence 2, Appli
21	411	16.8	7257	US-09-335-409-5	Sequence 5, Appli
22	411	16.8	7257	US-09-568-102-5	Sequence 5, Appli
23	411	16.8	7257	US-09-567-969-5	Sequence 5, Appli
24	411	16.8	7257	US-09-568-480-5	Sequence 5, Appli
25	411	16.8	7257	US-09-568-486-5	Sequence 5, Appli
26	411	16.8	7257	US-09-568-472-5	Sequence 5, Appli
27	411	16.8	7257	US-09-567-899-5	Sequence 5, Appli

28	409	16.7	3665	2	US-08-222-617A-13	Sequence 13, Appli
29	409	16.7	3712	2	US-08-222-617A-4	Sequence 4, Appli
30	409	16.7	3712	2	US-08-222-617A-25	Sequence 25, Appli
31	409	16.7	5032	4	US-09-538-092-979	Sequence 979, App
32	408	16.6	3461	3	US-09-334-220-2	Sequence 2, Appli
33	406	16.6	3969	3	US-08-061-376-5	Sequence 5, Appli
34	405	16.5	3869	4	US-09-538-092-1262	Sequence 1262, Ap
35	405	16.5	4861	4	US-09-919-497-70	Sequence 70, Appli
36	403	16.4	4866	4	US-09-424-783-2	Sequence 2, Appli
37	402	16.4	3460	3	US-09-334-220-1	Sequence 1, Appli
38	401	16.4	5405	3	US-08-718-388-9	Sequence 9, Appli
39	396	16.2	3959	2	US-08-970-269A-30	Sequence 30, Appli
40	386	16.2	3959	3	US-09-407-562-30	Sequence 30, Appli
41	394	16.1	5588	3	US-09-036-987A-6	Sequence 6, Appli
42	394	16.1	5588	3	US-09-370-700-6	Sequence 6, Appli
43	389	15.9	3443	2	US-08-416-603-2	Sequence 2, Appli
44	389	15.9	6396	4	US-09-410-551B-72	Sequence 72, Appli

## ALIGNMENTS

### RESULT 1

US-09-627-986-1  
; Sequence 1, Application US/09627986  
; Patent No. 6569616  
; GENERAL INFORMATION:  
; APPLICANT: Compton, Teresa  
; APPLICANT: Huber, Mary T.  
; TITLE OF INVENTION: HUMAN CYTOMEGALOVIRUS GLYCOPROTEIN O AS A NEW DRUG  
; FILE REFERENCE: 960296.96579  
; CURRENT APPLICATION NUMBER: US/09/627,986  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 60/146,180  
; PRIOR FILING DATE: 1999-07-29  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Human cytomegalovirus  
US-09-627-986-1

Query Match		98.7%;	Score 2420;	DB 4;	Length 465;
Best Local Similarity		99.4%;	Pred. No. 4.3e-72;		
Matches 463;		Conservative 0;	Mismatches 1;	Indels 2;	Gaps 2;
QY	1	MGRKEMVDRVPMFVLSISFLLVSFINCKVMSKALYNRPWGLVLSKIGKYKLDQLKL	60		
DB	1	MGRKEMVDRVPMFVLSISFLLVSFINCKVMSKALYNRPWGLVLSKIGKYKLDQLKL	60		
QY	61	EILRQLETTISTKYNVSKOPVKNLTNTTFFPQYIILAGPIQNYISITLWDFPSTQLRKP	120		
DB	61	EILRQLETTISTKYNVSKOPVKNLTNTTFFPQYIILAGPIQNYISITLWDFPSTQLRKP	120		
QY	121	AKVYVSOYNHNTAKTITFR-PPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMFF	179		
DB	121	AKVYVSOYNHNTAKTITFRPPCPGVRPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMFF	180		
QY	180	NVPRWNTKLYVGTQKYNVDSQTIYIFLGLTALLRYAQRNCTHSFYLVNAMSRLNRPVKY	239		
DB	181	NVPRWNTKLYVGTQKYNVDSQTIYIFLGLTALLRYAQRNCTHSFYLVNAMSRLNRPVKY	240		
QY	240	INGTKLNTMRKLKRQAPVKQFEKAKKTOSTTTPYFSYTTSAALNTVNTYITTA	299		
DB	241	INGTKLNTMRKLKRQAPVKQFEKAKKTOSTTTPYFSYTTSAALNTVNTYITTA	300		
QY	300	ARRVSTSTIAYRPDSSFMKSIATQRLDRTATVYITLRYQRNPFCEPFSNRRTAVSEFMKN	359		
DB	301	ARRVSTSTIAYRPDSSFMKSIATQRLDRTATVYITLRY-QNPFCEPFSNRRTAVSEFMKN	359		

QY 360 THVLIRNETPTIYGTLDMSLYNETMVENKTASDSNKTTPSPSMGFQRTFIDPLWD 419  
 Db 360 THVLIRNETPTIYGTLDMSLYNETMVENKTASDSNKTTPSPSMGFQRTFIDPLWD 419  
 QY 420 YLDSLLFLDEIRNFSLSRPTVNLTPPEHRRVAVNLSTNSLWMLQ 465  
 Db 420 YLDSLLFLDEIRNFSLSRPTVNLTPPEHRRVAVNLSTNSLWMLQ 465

RESULT 2

US-08-471-119A-2  
 ; Sequence 2, Application US/08471119A  
 ; Patent No. 5827706  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Leitner, Ernst  
 ; APPLICANT: Schneider, Elisabeth  
 ; APPLICANT: Schoergerdorfer, Kurt  
 ; APPLICANT: Weber, Gerhard  
 ; TITLE OF INVENTION: Cyclosporin Synthetase  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 5827706artis Corporation  
 ; STREET: 59 Route 10  
 ; CITY: East Hanover  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07936  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: 06-JUN-1995  
 ; APPLICATION NUMBER: US/08/471,119A  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kaasenoff, Melvyn  
 ; REGISTRATION NUMBER: 26,389  
 ; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201 503 8474  
 ; TELEFAX: 201 503 8807  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 15281 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Tolypocladium niveum  
 ; STRAIN: ATCC 34921  
 ; US-08-471-119A-2

Query Match 23.5%; Score 575; DB 2; Length 15281;  
 Best Local Similarity 1.7%; Pred. No. 2.2e-06;  
 Matches 254; Conservative 78; Mismatches 128; Indels 14673; Gaps 82;

QY 1 MG----- 2  
 Db 1 MGAIGQDMAYDLRANPSRASSISSNRYSEPVEQSPAQGLWFLHQLKLGASWDITPAAIR 60  
 QY 3 ----- 2  
 Db 61 LRGLHDIDALNAASRALTORHETLRTTFEQDGVGVQVHVASGLERGLRIVDASSRDLAQ 120  
 QY 3 ----- 2  
 Db 121 LLAEEQTMKFDLSEPAWRVALLKVAEDHHILSIWVHHIISDRSLDIIQQELGELYTAA 180

QY 3 ----- 2  
 Db 181 SQKKSISACPLGPPIQYRDLTWQNOQVAEQERQGYWIBQLDNNTPAELLTELPRP 240  
 QY 3 -----RKEMMV----- 8  
 Db 241 AIPSGETGKISFQIDGSHVHELLAFCRSQQVYAYVLLAAFRVAFHFLTGAEDATIGAPV 300  
 QY 9 -----RDVP----- 12  
 Db 301 ANRDPELENVAVPLATLQCMRVVLDDDTFESVLRLQIMSVMTAEAHNRDVFERIVSAL 360  
 QY 13 -----KMFVL----- 17  
 Db 361 LPGSTDTSRHPLVQLMFALHPAODTGRARWGFLAEETLQSAAPTRFDMEMHLFEGDDRFD 420  
 QY 18 ----- 17  
 Db 421 ANVLFTGTLFDABAIRSVSVSIFREVLRRGISSEPAHVHVKTMPLTDGLAAIRDMGLLDIGTT 480  
 QY 18 ----- 17  
 Db 481 DYPREASVVDMEQEQVALNPSATAVADASSRLSYSELDHKSQDLAAWLRRLRQLKPTLIG 540  
 QY 18 ----- 17  
 Db 541 VLSPPSCETWVSFLGILKAHLAYLPDLINVPLARIESILSAVDGHKLVLLGSNVPOPKVD 600  
 QY 18 ----- 17  
 Db 601 VPDVELLRISDALNGSQVGLAGKQATAPKPSATDLAVIFTSGSTGKPKGVMIHRGIVR 660  
 QY 18 ----- 17  
 Db 661 LVKGTNIISPAQAAVPTAHLANTAFDLSTWEIYTPILNGTTLVCIHESVTLDSKALEAVF 720  
 QY 18 ----ISISFL----- 23  
 Db 721 TKEGIRVAFAPALIKQCLADRPAPFAGLDSLYAIGDRDRDALHAKSLVKHGVYNAYG 780  
 QY 24 ----- 23  
 Db 781 PTENSVVSTIYSVSEASPVPTGVVGRASNGAYVMDQDQQLVSPGVMGELVVGDLA 840  
 QY 24 ----- 23  
 Db 841 RGYTDSALDKNRFVVVQIDGESIRGYRTGDRARYSLKGGQIEPFGMDQQVKIRGHRIEP 900  
 QY 24 ----- 23  
 Db 901 AEVEHALLNSQVRDAAVVIRRQEESEEPAMIAFVTTQGTLPDLHVNINGNHGVPDNGSK 960  
 QY 24 ----- 23  
 Db 961 NDQFAVHVESELRRRLQMLLPSYMPARIIVLDHPLNPNKGVDRKALQSQAKTVQSKL 1020  
 QY 24 ----- 23  
 Db 1021 VSRVAPRNEIEAVLCEYSYVLGVEVGITDNPFDLGGHSLTAMKLAARISQRLDIQASV 1080  
 QY 24 ----- 23  
 Db 1081 ATVPEQPLADLAATTQRGSTLYSVIPTTEYTPVEQSPAQGLWFLHQLTGSWNVNM 1140  
 QY 24 ----- 23  
 Db 1141 LTVRLRGLHDVLDALGTALLALEKRHETLRTTFEERDGVGMQVHSHSLMGLRLIDISEKS 1200  
 QY 24 ----- 23  
 Db 1201 GTAAHEALMKEQSTFRDITREFGWRVALLKLADHHIFSIWVHHIVSDGSLDLRLHQLG 1260  
 QY 24 ----- 23

Db 1261 LYSALRGDPLSRLEPLIQYRDFAVWQKDSQKAAHQRLQLEYWTKQLADSTPAELLT 1320  
QY 24 ----- 23  
Db 1321 DFRPRLSGKAGKVPVAIEGSLYDTLQVSRTHQVTSFAVLLAAFRAAHFRLTGSDNAT 1380  
QY 24 ----- 23  
Db 1381 IGVSANRRPELENVIGFVNTQICIRITIDENDFESLVRQVRSTTTAAQNDQVPFEQ 1440  
QY 24 ----- 23  
Db 1441 VVSSLMPSSSRDASRNPLVOLMFALHGQDLFKIQLEGTEBEEVPTTEVTRFDIEPHLYQ 1500  
QY 24 ----- 23  
Db 1501 GASKSGDIIPAADLPEAETIRGVSVFQEVLRGLQQPOTPIWMTPLTDGIPELERMGL 1560  
QY 24 ----- 23  
Db 1561 LHMVKTDYPRNMSVDVFOQQVRLSABEATAVIDSSRMSYAELOQSDQVAAMLROQLP 1620  
QY 24 ----- 23  
Db 1621 AETFVAVLAPRSCAVIALFGILKAGHAYLPLDVNPAARLAILABVKGKLVLLGAGE 1680  
QY 24 ----- 23  
Db 1681 PSEGOQPEVSIURIADATSPAGHASLRDCKSKPTAGSLAYVIFTSGSTCKPKGVMEHR 1740  
QY 24 ----- 23  
Db 1741 GVLRVKTNILSSLPPAQTFRMAHNSLAFDASIEVFTALENGGSLVCIDRFTILDAQ 1800  
QY 24 ----- 23  
Db 1801 ALEALFLREHINIALPPALLKQCLTDAATIKSLDLYVGGDRDLPDADAALAKALVKSE 1860  
QY 24 ----- 23  
Db 1861 VYNAYGPTENTVMSTLYSIADTERFVNGVPIGRAVNSGVYVMDQNLVPLGVMSGLVV 1920  
QY 24 ----- 23  
Db 1921 TGDGLARGYTNPALDSRFDVVIARGQLLRAYRTGDRARYRPKQGVQVEFFGRMDHQVKVR 1980  
QY 24 ----- 28  
Db 1981 GHRIELAEVEHALLSSAGVHDVVVNSQBDNQGVENVAFITAQDNETLQEAQSNQVQE 2040  
QY 29 ----- 28  
Db 2041 WESHFETAYADITADONTLGRDFTSWTSMYDGLIDKREMQEWLDDTMRFTLDGQAAG 2100  
QY 29 ----- 28  
Db 2101 HVLEIGTGTGWLNLGQAGLKSIVIGLEPSQSAVFVNKAAQTFPGLEGKAQVHVGTAMD 2160  
QY 29 ----- 28  
Db 2161 TGRLSALSPDLIVINSVAQYFPPSREYLAEVVEALVRIPGVRRIFFGDMRTYATHKDFLVA 2220  
QY 29 ----- 28  
Db 2221 RAVHTNGSKVTRSKVQOEVARLEEBEELLVDPAFTLSKESLSEBIEHVLEILPKMKVN 2280  
QY 29 ----- 28  
Db 2281 NELSSRYGAVLHRNHQNSRSIHKINAESWIDPASSQMDRQGLARLLKENKDAESIA 2340  
QY 29 ----- 28

Db 2341 VFNIPIYSKTIIVERHIAKSLADDDHDDTHSSIDGVAMISAAREKASQCPSLDVHDLVOLA 2400  
QY 29 ----- 28  
Db 2401 EDAGPRVEVSUARQSRONGALDVFFHHFQPTENESRALVDFPTDYKGOQARSLTNRPLQR 2460  
QY 29 ----- 28  
Db 2461 VESRRIEAQVREQLQVLLPAYMIPARIVVLQNNPLNTSGKVDKRELTLRKAVTAARTPSS 2520  
QY 29 ----- 36  
Db 2521 ELVAPRDSIEAIIKCFKDVLGVEVGITDNFNVGSHSLATYLAARLSQNLQAIAVKD 2580  
QY 37 LYNRP----- 41  
Db 2581 IFDRPVIADLAATIQQDTTTEHNPILPSTYGPVEQSPAQGLWFLDQLNVGATWYLMPPA 2640  
QY 42 ----- 41  
Db 2641 VRLRGPLVVSALAAALLALEERHETLRTTFIEQEGIMQVIHPFAPKELRVIDVSGEES 2700  
QY 42 ----- 41  
Db 2701 TIQKILEKEQTTFFNLASEPGRFLALLKTGEDEHILSTVWHHAISDGWSVDIFQOEIGQF 2760  
QY 42 ----- 43  
Db 2761 YSAILRGHDPLAQIAPLSIQYRDPATWQRIQFOVAEHRQLAYVTWKQLADNKPALLTDF 2820  
QY 44 ----- 49  
Db 2821 KRPPMLSGRAGEIPVVVDGLIYEKLODFCIRIQVTAFTVLLAAFRAAHYRMTGTEDATIG 2880  
QY 50 ----- 49  
Db 2881 TPIANRRPELEGLIGFVNTQCMRITVDVESFETLVHQVRETTLAHANQDVPEQIV 2940  
QY 50 ----- 69  
Db 2941 SNILPGSSDTSRNPLVQMFALHSQQNLGKVRLEGIEEBIIISIAETTRPDIEPHLYQEA 3000  
QY 70 ----- 69  
Db 3001 RLNGSIVYAADLFPVETIQSVITIFQGIQKGLGEPDMPVSMALDGLSLSSTGLLHP 3060  
QY 70 ----- 69  
Db 3061 QOTDYPCDASVVQIFKQVAVNPVDVIAVRDESTRLSYADLDRKSDQVACWLSRRGIAPET 3120  
QY 70 ----- 69  
Db 3121 FVAILAPRSCETTIVAILGVLKANLAYPLDVNVNVPASLEAILSEVSGSMLVLVGAETPIP 3180  
QY 70 ----- 69  
Db 3181 EGMAEATIRITIELADAKTDINGLAASQTAASLAYVIFTSGSTRPKGVNVHRGIV 3240  
QY 70 ----- 69  
Db 3241 RLTKQNTSKLPESPHMAHINLAPDASVWEVFTLLNGGTLVCIDYFTLLESTALEKV 3300  
QY 70 ----- 69  
Db 3301 PFDQVNVALLPPALLKQCLDNSPALVKTLVLYIGDRLDASDAKARGLVQTQAFNAY 3360  
QY 70 ----- 80  
Db 3361 GPTENTVMSTIPIAEDDPFINGVPIGHAVNSGAFVMDQNOQITPPGAMGELIVTGDGLA 3420  
QY 81 ----- 80  
Db 3421 RGYTTSSLNTRGFINVNDIGEQVARYRGTGDRVRYRPKDLOIEFFGRIDHQVKIRGHRIEP 3480

QY	81	-----	130	-----	QY	130	-----	129
Db	3481	AEVEYALLSHDLVTDAVTHSQENODLEWVGFVAARVADVREDESSNQVQEWQTHFDSI	3540	-----	Db	4561	FHEQVSNPDSIALIHGSEKLSYAQLORESDRVARWLRHRSFSDTLIAVLAPRSCETII	4620
QY	81	-----	80	-----	QY	130	-----	129
Db	3541	AYADITTIQQSLGRDPMFSWTSMYDGLIKKSQMOWELDDTMRSLDSQPPGHVLEVTG	3600	-----	Db	4621	APLGLKANLAYLPDVKAPAARIDAIVSSLPGNKLILLGANVTPPKLQEAADFPVIRD	4680
QY	81	-----	80	-----	QY	130	-----	129
Db	3601	TGMVFNLEGGELQSYVCLSPSATAFVNKAASKFPLEDRIRVEVGATDIDRLGDD	3660	-----	Db	4681	TFTTLTDGTLQDGPTIERPSAQSLAYAMFTSGSTGRPKGMVQVHRNIVRLVKNNSVNAKQ	4740
QY	81	-----	85	-----	QY	130	-----	129
Db	3661	LHAGLVVNSVAQYFSPQDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDLVARAVHAL	3720	-----	Db	4741	PAARIAHISNLAFDASSWEIYAPLLNGAIVCADIYFTTIDPOALQETFOBEHIEIRGMLP	4800
QY	86	-----	85	-----	QY	130	-----	129
Db	3721	GDKATKAEIQREVVRMESEDELIVDPAPFTSLTQVENIKHVEIILPKRMATNELSSYR	3780	-----	Db	4801	PSLLKQCLVQAPDMISRLDILFAAGDRFSSVDALQAQRLVGSVFNAYGPTENTILSTIY	4860
QY	86	-----	85	-----	QY	130	-----	129
Db	3781	YAAVLHVNDLAKPAHKVSPGAWVDFAATKMDRDALLRLRGTKISDHIAIANIPNSKTIV	3840	-----	Db	4861	NVAENDSFVNGVPIGSAVNSGAYIMDKNQQLVPAGVMGELVWTGDGLARGYMDPKLDAD	4920
QY	86	-----	85	-----	QY	130	-----	139
Db	3841	ERTICESVYDLGGDAKDNDRVSWLSAARSNAVKNASLAIDLVDIAQAGPRVEISCAR	3900	-----	Db	4921	RFIQLTVNGSEQVRAVYRTGDRVYRPKDFQIEFFGRMDQOIKIRGHRIEPAEVEQAFLND	4980
QY	86	-----	85	-----	QY	140	-----	139
Db	3901	QMSONGALDAVPHLGPSPQSSHVLIDFLTDHQRPEEALTNWHLRAQSRVERQIRER	3960	-----	Db	4981	GFVEDVAIVIRTPENQEPENVAFTVAKGNSARBEZATTOEGWEAHFEGGANTEEIE	5040
QY	86	-----	97	-----	QY	140	-----	144
Db	3961	LQTLIPAYMIPAQIMVLDKPLNANGKVRKQLTQRAQVTPKAKQVAPVAPTEIERVL	4020	-----	Db	5041	SEALGYDFMGWTSVMYDGEIDKDEMRWINDTMRSLDGRKGRVLEVTGTGIMFNLG	5100
QY	98	-----	97	-----	QY	145	-----	144
Db	4021	COBFSDVLGVDIGIMENFDLGGHSLMATKLAARISRLETHSVKGEIFDHPRVCDLVI	4080	-----	Db	5101	RSQGLERYIGLEPAPSAAEFVNNAKSFPLAGRAEVHVHTAADVGTLOGLTSDMAVINS	5160
QY	98	-----	110	-----	QY	145	-----	148
Db	4081	VQGSAPHPDPIVSTKYTGVFQSFQAQGLWFLDQNFCAWYMLPLAVRLGAMNVHALT	4140	-----	Db	5161	VAQYFPTPEYLAETIKSLVQVPCMKIYILGDMRSHAMNRDFAAARAAYSLADNASKDRVR	5220
QY	111	-----	110	-----	QY	149	-----	155
Db	4141	AALLALERRHELLRTTFYEQNGVMQKVPVVTETLRIIDLNGDGDYLPFLKKEQTAPF	4200	-----	Db	5221	QKMEELEKEEBELLVDPAPFTALASQLODRIQHVETLPRKMKATNELSSYRYAAVLHISD	5280
QY	111	-----	117	-----	QY	156	-----	165
Db	4201	HLETEPGWRVALLRGPDYILSVVMHHIISDQMSVDVLFQELGQFYSTAVKGDHPLSQT	4260	-----	Db	5281	EPLFIYKIDPEAWINFEGRSLTREALAQVLKENENAESVAISNIPYSKTVVVERHIVRSLD	5340
QY	118	-----	120	-----	QY	166	-----	165
Db	4261	TPLPIHYRDFALWOKKPTQBSHERQOYVWEQVDSAPAEILLTDLPRPSILSGQAGEMS	4320	-----	Db	5341	QEDANAPESMGSDWISAVRTRAQQCHTSLASDLFDIAEDAGFRVSVWARQHSQHAL	5400
QY	121	-----	123	-----	QY	166	-----	167
Db	4321	VTIEGALYKNLEBFCEVRHVTSFVLLAALRAAHYRLTGSEDATIGTPIANRNPPELQI	4380	-----	Db	5401	DAVFHHLKPAVEDSRVLKFPETHQGRPLKSLTNQPLLPQASRRRAELLIREGLQTLPPY	5460
QY	124	-----	123	-----	QY	168	-----	172
Db	4381	IGFFVNTQCIKITVNEDETFSVLQVQVRSTATAFAHQDVPEKIVSTVLLPGSRDARNP	4440	-----	Db	5461	MTPSQITLIDRMPNLNANGKVDRELRARRAKITOKSKPVEDIVPPRNSVEATVCKGTFDVL	5520
QY	124	-----	129	-----	QY	173	-----	181
Db	4441	LVQLMAVHSHQKXLBKLENASHVVPTEITRFDLEFHLFQDDKLEGSILYSTDLFE	4500	-----	Db	5521	GVEVGITDN--FFNLGCHSLMATKLAARLGRQLNTRISVRDVPDQPVADLAIVQRNSA	5578
QY	130	-----	129	-----	QY	182	-----	181
Db	4501	AVSVOSLLSVFOEILRLRGLNGDPVPISTLPLQDGIQVLDQRLGDLVQKTEYPRDSSVVDV	4560	-----	Db	5579	PHEPIKADYTGVPVPOSFAQRLWFLDQNLVGTWYMLPLGIRLHGLSLRVDALATAISAL	5638
					QY	182	-----	181

Db	5639	EQRHEPLRTTFHEEDGVGVQVQDHRPKOLRIIDLSTQPKDAYLVLKHEQTTFLDLATE	5698	Db	6718	IYELEANEELLTDPAPFTSLRTRLGEKIKHVEILLPKTKATNELSKYRYAAVLHVRGSR	6777
Qy	182	PRWNTKL-----	188	Qy	211	-----	210
Db	5699	PGWRVALIRLGBEEHILSIVMHHIISDGWSVEVLFDEMHRFYLSALRQDDPMEQILPLPI	5758	Db	6778	EQSTIHQVSPNAWIDFADGLDRQTLINLLKEHKDAGTVAIGNIPYSKTIVERFVNKSLS	6837
Qy	189	-----	188	Qy	211	-----LLRYAQRNCT-----	220
Db	5759	QYRDFAAWQKTERQVAHQDQDYWEHLADSTPAELLTDLPRPSILSGRANELPLTIEG	5818	Db	6838	EDDMEEQNSLPGSAWVAARMAOQSPSLDAMDVKIEAQEAGYQVEVSWARQWONGAL	6897
Qy	189	-----	188	Qy	221	-----	220
Db	5819	RLHDKLRAFRCVRHOATPFVILLAAALRAAHYRLTGASDATLGTPIANRNPPELENMIGFFV	5878	Db	6898	DAIFHFEPKEGARTLIEFPTDYEGRNVTLTNRPLNSIQSRRLGTQIREKLQTLPPY	6957
Qy	189	-----	188	Qy	221	-----	220
Db	5879	NTQCMRIATENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQVI	5938	Db	6958	MIPSRIMVLQMPVNNNGKIDRKELVRAIVAPKPRSAATRVAPRNEIEAILRDEPEDVL	7017
Qy	189	-----	188	Qy	221	-----HS-----	222
Db	5939	LAVHSQQDLGKLTLEGLRDEAVDSATSTRPDVEFHLPEHADRLSGSVLYAKELFKLRTIE	5998	Db	7018	GTEVSVLDNFPDLGGHSLMATKLAARVSRRLDAHISIKOVFQDQVLADLAASIQRSEAPH	7077
Qy	189	-----	188	Qy	223	-----	222
Db	5999	SVVSFLETLRALDQPLTPLAVPLTDGVGEIASKGLLDVPRDYPRDANIIVEVFQHV	6058	Db	7078	EPQPYPYTGPAEQSPAQGRLWFLDQNLGATWYLMPLAIRRGQLRVAALSAAALFALER	7137
Qy	189	-----	188	Qy	223	-----	222
Db	6059	RATPDAIVKDATSIITYAQDQSDRLAIWLSRRHMPETLVGLAPRSCETIIAMFGI	6118	Db	7138	RHETLRTTTEESDGVGVQIVGEARNSDLRVHDVSTGDDGEYLEVLRREQTVFPDLSEPG	7197
Qy	189	-----	188	Qy	223	-----	222
Db	6119	MKANLAYPLDINSPAARLSILSAVDGNKLVLLGSGVTAPEQENPEVEAVGIOEILAGT	6178	Db	7198	WRVCLVKTGEEDHVLISVMHHIYDGSVDYILRGLGQFYSAALRGQDPLLANPLPIQY	7257
Qy	189	-----YV-----	190	Qy	223	-----	222
Db	6179	GLDKQGSNARPSATSLAYVIFTSGTGPKGVMEHRSVTRLAKPSNVISKLPOGARVA	6238	Db	7258	RDFAAWQREAKQVEEHQORQLGYWSKQVLDSTPAELLTDLPRPSILSGRAGSVDTTIEGVS	7317
Qy	191	-----	190	Qy	223	-----	222
Db	6239	HLANIAFDASIWEIATLLNGATLVCLDYHTVLDLDCRTLKEVPERESITVVTLMPALLKQC	6298	Db	7318	YGALQSFCTRSTVTFVLLTVFRIAHFRLTAVDATIGTPIANRNPPELTLVGCFVNT	7377
Qy	191	-----	203	Qy	223	-----	222
Db	6299	VASIPETLAHLDLLYTGDRVGHDAMARSLVKIGMFGSGYGTENTVIS-TIYEVDAD	6357	Db	7378	QCMRISIADDNDFEGLVRQVRNVATAAYANQDVPFERIVSALVPGSRNTRSNPLVQLMFA	7437
Qy	204	-----	203	Qy	223	-----FVLV-----	236
Db	6358	MFYNGVPIGKTVNSGAYVMDRNQQLVPSGVVGGELVVTGDGLARGYTDPSLNKNRFIYT	6417	Db	7438	VOSVEDYDQVRLEGLSVNMPGEASTRFDMFHLVPGDQKLTGSLVLYSSDLFQGTIQNP	7497
Qy	204	-----	203	Qy	227	-----NAMSR-----	235
Db	6418	VNGESIRAYRTGDRVRYRPHDLQIEFPFGMDQVKIRGHRIBEPGEVESALLSHNSVQDAA	6477	Db	7498	VDIFQECRLSVDQPLTPIISVLPPFSNAISNLESLDLLEMTSDYPRDRTVVDLFRQAAI	7557
Qy	204	-----	203	Qy	236	-----	235
Db	6478	VVICAPADQDSGAEMVAFVAARNTEDEDTQEEBAVDQVQGWETHFETAAYSEVKDIRQSE	6537	Db	7558	CPDSIAVKDSSQLTYAQLDEQSDRVAAWLHERHMPAESLVGVLSPRSCETIIAYFGIMK	7617
Qy	204	-----FLGLTAL-----	210	Qy	236	-----VPK-----	238
Db	6538	VGNDFMGWTSMDGSEIDKTMHEWLNDRMTILDAREPHGVLEIGTGTGMVFNFLAKCP	6597	Db	7618	ANLAYPLDVYAPDARLAAIILDTVEGERLLLLGAGVPQPGIQIPRLSTAYIAEALSHATT	7677
Qy	211	-----	210	Qy	239	-----	238
Db	6598	GLOQYVGFEPKSAAPFNDAASFPALKDGRSIVHVGTATDINKAGPIQPRLVVINSVA	6657	Db	7678	VDVTSIPQPSATSLAVIIFTSGSTGPKGVMIHHRGIVRLVRDTRNVNVPFESSALPVSH	7737
Qy	211	-----	210	Qy	239	-----	238
Db	6658	QYFPTPEYLFRVVEALVQIPFSVERIVFGDMRTNAINRDFVASRALHTLGEKANKRLVRQM	6717	Db	7738	FSNLAMDAATWEIYTAVLNGGTVCIDRDTMLDIAALNSTFRKENVRAAAFPTPAFLKQCL	7797
Qy	211	-----	210	Qy	239	-----	239
				Db	7798	ATPELVLVANLEILHTAGDRDLDPGDANLACKTAGKGIENVLGHNTENTAYSTFYFVVGEEPT	7857

Qy	240	ING-----	242	Qy	280	-----YTTSA-	284
Db	7858	NGVPGVGRGISHAYIIDRHQKLVAGVMGELITGDGVARGYTDALANKDRFVYIDIN	7917	Db	8938	GNQOLVSPGVMEIIVVTGDLARGYITDSALDEDRFVHVITIDGEENIKAYRTGDRVYRPK	8997
Qy	243	-----	242	Qy	285	-----	284
Db	7918	GKSTWSYRTGDKARYRPRDQGLEFFGRMDQWKIRGVRIEPEGEVELTLLDHKSVLAAATVV	7977	Db	8998	DFEIEFFGRMDQOVKIRGHRIBPAEVEHALLGHDLVHDAVAVLKRPANQPEMIAFITSQ	9057
Qy	243	-----	242	Qy	285	-----	284
Db	7978	VRRPNDGPEMIAFITDAEDDVQTHKAIYKHLOGILPAYMIPSHLVILDMPVTNGKV	8037	Db	9058	EDETIEQHESNKQOVQWGEHFDVSRYADIKDLDTSTFGHDFLGWTSMTYDGVDPVNMEMKE	9117
Qy	243	-----	242	Qy	285	-----	284
Db	8038	DRKDLALRAQTVQKRRSTAAARVPRDEVAEVLCEBSNLEVEVGITDGFDDLGHSHLLA	8097	Db	9118	WLDETTASLLDNRPFGHILEIGAGTCMILSNLKGVDGLQKYVGLDPAPSAALFVNEAVKS	9177
Qy	243	TKLK-----NT-----	248	Qy	285	-----	284
Db	8098	TKLAARLSRQLNTRSVKVDQFIPILADLADIIIRRGSHRDPPIPATPYTGVEQSFQAGR	8157	Db	9178	LPFLAGKARVLGTALDIGSLDKNEIQPELVVINSVAQYFPTSEVYLKVVKAIVEVPSVK	9237
Qy	249	-----	248	Qy	285	-----ALN-----	287
Db	8158	LWFLEQLNGASWYLPFAIEMRGPIQTKALAVNALVHRHEALRTTFEDHGVGVQVI	8217	Db	9238	RVFFGDIRSQALNRDPLAARAVRALGDNASKSQIREKIAELSEEBELLVDPAFFVSLRS	9297
Qy	249	-----MRKLKXKQA-----	257	Qy	288	-----	287
Db	8218	QPKSSQDLRIIDLSDAVDDTAYLAALKREQTAFDLTSEPGWRVSLRLGLDDYILSIVM	8277	Db	9298	QLPNIKHVEVLPKMKATNLSYRYAAVLHISHNEEBEQLLIQIDIDPTAWVDFAAATQKDS	9357
Qy	258	-----PVKEQ-----	262	Qy	288	-----VTTNVTYSITTAARRV-----	303
Db	8278	HHIISDGTVDLRLQELGQFYSAAIRQBEPLSQAQSLPIQYRDFAVMQROENIQEQAQK	8337	Db	9358	QGLRNLOQGRDDVMIAVGNIPYSKTIVERHIMNSLDQHVNSLDGTSWISDARSAAATC	9417
Qy	263	-----	262	Qy	304	-----	303
Db	8338	LKYSQQLADSTPCEFLTDLPRPSILSGEADAVPMVIDTVYQLLTDPCRTHQVTSFVUL	8397	Db	9418	TSFDAPALTOLAKEGFRVELSWARQSONGALDAVHRLATDANCERSRVLVHPTDHO	9477
Qy	263	-----FEKKAKK	269	Qy	304	-----	303
Db	8398	LAAPRTAHYRLTGTLDTATVGTPIANRNRPELEGLIGFVNTQCMRAISSETFESLVQQ	8457	Db	9478	GRQLRTLNRPLQRAQSRRIESQVFEALQALPAYMIPSIIVLPQMPNTNANGKVDKQOL	9537
Qy	270	TQSTTTTPYFS-----	279	Qy	304	-----	303
Db	8458	VRUTTEAFANQDVPPEQIVSTLLPGSRDTSRNPVLQVMPALQSQDQLGRIOLEGMTDEA	8517	Db	9538	ARRAQVAKRKAVSARVAPRNDTEIVLCEEYADILGTEVIGITDNFDFMGHSLMATKLAA	9597
Qy	280	-----	279	Qy	304	-----	303
Db	8518	LETPLSTRLDLEVHLFQEVGKLSGLLYSTDLEFEVETIRGIVDVDFLEILLRGLQPKQRL	8577	Db	9598	RLSRRLDTRVTYKVEYFDKPEVLADLAASIEQGSTPHLPIASSVYSGPVEQSYAQGRLWFLD	9657
Qy	280	-----	279	Qy	304	-----	303
Db	8578	MAMPITDGITKLRDQGLLTVAKPAYPRESSVIDLFRQVAAAAPDAIAVWDSSTLTYYADL	8637	Db	9658	QFNLNATWYHMSLAMRLLGLPLNMDALDVALRALEQRHETLRTTTPBAQKDIGVQVVEAGM	9717
Qy	280	-----	279	Qy	304	-----	303
Db	8638	DGQSNKLAHWLCORNMAPETLVAVFAPRSLCTIVAFLGVKANLAYPLDVNAPARIAEA	8697	Db	9718	KELKVLDSLKNEKEHMAVLENEQMRPFTLASEPGWKGLHARLGLTPTTEYILSLVMMHMFSD	9777
Qy	280	-----	279	Qy	304	-----	303
Db	8698	ILSAVPGHKLVLVQAHGPELGLTMADELVLQIDEALASSSGDHEQIHASGPTATSLAV	8757	Db	9778	GWSVDILRLQELGQFYSAALRGDRDPLSQVKPLPIQYRDFAAWQKEAAQVAEHERQLAYWEN	9837
Qy	280	-----	279	Qy	304	-----	303
Db	8758	MFTSGTGPKGMIDHRSIIRLVKNSDVVATLPTFVRMANVSNLAFDISVQBIYITALIN	8817	Db	9838	QLADSTPGELLTDFPRPQFLSKAGVPIPVITIEGPVVEKLLKFSKQVTLFSLVLLTAFRA	9897
Qy	280	-----	279	Qy	304	-----	307
Db	8818	GGTLVCLDYLTLLDSKILYNVFEAQVNAAMFTPLVLLKQCLGNMPAIIISRLSVLFNVGDR	8877	Db	9898	THFRLTGAEDATIGTPIANRNRPELEHIIIGFPVNTQCMRLLLDTCGSTFESLVQHVRSVAT	9957
Qy	280	-----	279	Qy	308	IAY-----	310
Db	8878	LOADAVAASGLIQDAVYNAYGPTENGMSQTMVKVDVNEPFFVNGVFIGRSITNSGAYVMD	8937	Db	9958	DAYSNOQIPFERIVSALLPGSRDASRSLIQLMFALHSQPDIGNITLLEGLERLPTSPA	10017
				Qy	311	-----	310



Db	10018	TRFDMFHLFQEPNKLSGSILFADELFPQPETINSVVTVFOEILRGLDQPVQSISTMPLT	10077
QY	311	-----	310
Db	10078	DGLIDLEKLGLESSNPPRDYSVDVFRQVVAANPNAPAVVDSETSMTSLDQKSEQ	10137
QY	311	-----RPDS-----	314
Db	10138	IAAWLHAQGLRPESLTCVMAPRSFETIVSLFGILKAGAYPLDVNPAARIQIPILSEVE	10197
QY	315	-----	314
Db	10198	GKRLVLLGSGIDMPQSDRMDVETARIQDILNTNKTVERSDPMSRPSATSLAYVFTSGTG	10257
QY	315	-----	314
Db	10258	RPKGVMIEHRNIIRLVKQSNVTSQPLQDLRMAHISNLPDASIWEIFTAILNGGALICID	10317
QY	315	-----SFMK-----	318
Db	10318	YFTLDSQALRTTFEKARVNATLFPALLKECLNHAPTLPFEDLKVLYIGDRLDATAAK	10377
QY	319	-----	318
Db	10378	IQALVRGTVNAYGPTENTVMSTIYRLTGESYANGVPIGNVSSGAYIMQKQRLVPP	10437
QY	319	-----	318
Db	10438	GVMGELVSGDLGARGYTNSLNADRPDIVINDQKARVYTGDRTRYRPKGSIEFFGR	10497
QY	319	-----	318
Db	10498	MDQQVKIRGHRVPAEVEQAMLGNKAIHDAAVVVQAVDQGETEMIGFVSNASDRPSEGBE	10557
QY	319	-----	318
Db	10558	EITNQVQWEDHFESTAYAGIEAIDQATLGRDFTSWTMYNGNLIDKAEMEELDDTMS	10617
QY	319	-----	318
Db	10618	LLDKEDARPCAEIGTGTGMVLFNLPRKNDGLESVVGIEPSRSALFVDKAAQPPGLOGKT	10677
QY	319	-----	318
Db	10678	QILVGTAEEDIKLVKDFHPDVVVINSVAQYPPRSYLVQIASELIHMTSVKTIFFGDMRSW	10737
QY	319	-----	318
Db	10738	ATNRDPLVSRALYTLGDKATKQIQROEVARLENEDELLVDPAPFTSLTSQWPGKVHVE	10797
QY	319	-----	318
Db	10798	ILPKRMRTSNELSSRYAAVLHICRDEGRNRYGRRVHSVEENAWIDFPASSGMDRHALVQ	10857
QY	319	-----	318
Db	10858	MLDERDAKTVAIGNIPHSNTINERHFTTSLDTEGEGIAQDSLDSAWQSATKAMAARCP	10917
QY	319	-----	318
Db	10918	CLSVTELVEITGAAGFRVEVSWARQSRQHGLDWVFHHLDDRDVGRVLINFPTDFERLPP	10977
QY	319	-----	318
Db	10978	STGLTSRPLQRIQNRPFESQIRESQLQTLPLPPYVPSRIVVLERMPLNANSKVDKELARK	11037
QY	319	-----SIMATQL-----	325
Db	11038	ARTLOTIKPSATRVAPRNDIEAVLCDFQAVLGVTVGVMDNFPFELGSHSLMATKLAARLS	11097
QY	326	-----	325

Db	11098	RRLDTRVSVKDI PNQPILODLADVVQTSAPHEAIPSTPYSGPVEQSFSGRLWFLDQLN	11157
QY	326	-----	325
Db	11158	LNASWTHMPLASRLRGPLRIEALQSALATIIEARHESLRTTPEEQDGPVQIVRAARNKQL	11217
QY	326	-----	325
Db	11218	RIIDVSGTEDAYLAALKQBQDAAFDLTABPGWRVALLRLGPDHVLISVMHHIISGWSV	11277
QY	326	-----RDLATW-----	331
Db	11278	DILRQELGQLYSNASSQPAPLPQVYRDPALWQKDSQIAEHQKQLNWKQRLVNSKPAEL	11337
QY	332	-----VYTLR-----	339
Db	11338	LADFTRPKALSGDADVIPETIDDDQVYONLRSFCRARHVTFSVALLAAPRAAHYRLTGAED	11397
QY	340	-----	339
Db	11398	ATIGSPIANRNPELEGLIGCFVNTQCLRI PVKSEDTPTLVKQARETATEAQDNQDVPP	11457
QY	340	-----QNPF-----	343
Db	11458	ERIVSSWVASSRDTSRNPLVQVMFAVHSQHDIGNIRLEGVEGKPVSMAASTRFDEAMHLF	11517
QY	344	-----	343
Db	11518	EDQMGNGNVVSKDLFESETIRSVAVFQETLRRGLANPHANLATPLTDGLPSLSLC	11577
QY	344	-----	343
Db	11578	LQVNOPYPRDASVIDVFREQVASIPKSIADIVASSQLTYTELDERSSQLATWLRQVTV	11637
QY	344	-----CE-----	345
Db	11638	PEELGVGLAPRSCETIIAFLGIKANLAYPLDVPNAPAGRIETILSSLPONRUILLGSDT	11697
QY	346	-----	345
Db	11698	QAVKLHANSVRTRISDALVESGSPTEELSTRPTAQSLAYVMTSGTGVKPVKVVHR	11757
QY	346	-----	345
Db	11758	GITRLVNSNVAKOPAAAAIAHLSNIAFDASSWEIYAPLLNGTGVVICIDYTTIDIKAL	11817
QY	346	-----	345
Db	11818	EAVFKQHHIRGAMLPALLKQCLVSAPTWISSLEILFAAGDRSSQDAILARRAVSGVY	11877
QY	346	-----	345
Db	11878	NAYGPTENTVLSITHIGENAEAFSGVPIGNVNSGAFVMDQNQQLVSAGVIGELVVVG	11937
QY	346	-----	345
Db	11938	DGLARGYTDKLRVDRFIYITLDGNRVRAYRTGDRVHRHPKQIQIEFFGFRMDQIQIRGH	11997
QY	346	-----	345
Db	11998	RIEPAEVEQALARDPAISDAVITQLTDEEBPELVAFPSLKNANNTGNGVSGVSDQEKID	12057
QY	346	-----PS-----RNRTAV-----	353
Db	12058	GDEQHALLMENKIRHNLOALLPTYMIPSRIIHVDQLPVNVANGKIDNELAVRAQATPTS	12117
QY	354	-----SEF-----	356
Db	12118	SVSTVAPRNDIETIICKEFADILSVRIGITDNFDLGGHSLIATKLAARLSRLDTRVS	12177
QY	357	-----	356
Db	12178	VRDVFDPVVGQAASIQQGSTPHEAIPALSHSGPVQOSFAQGRWLFLDRFNLAANYIM	12237

QY	357	-----MKNTHVLIR-----	365	QY	414	-----	413
Db	12238	PFGLRGRLRVDALQALRALBERHELLRTTFEBQDGVGMQIVHSPMRDICIVDDISGA	12297	Db	13318	GSNASKAMVRQOVAKLEDDEEELLVDPAFTLSLSDQFPDEIKHVEILPKRMATNELSSY	13377
QY	366	NE-----	367	QY	414	-----	413
Db	12298	NEDLAKKEEQAPPNLSTEVAMRVALKAGENHHILSIVMHHIISDGSVDIFIQBELAQ	12357	Db	13378	RYAAVHVGHQHPNGEDEDKQWAVKDINPKAWVDFAGTMDRQALLQLLODRQRQDDVV	13437
QY	368	-----	367	QY	414	-----	413
Db	12358	FYSVAVRGHDPLSQVKPLPIHVRFAVMQRQKVAVHESQLOXWIBQLADSTPABILSD	12417	Db	13438	AVSNIPYSKTIEMRHLQSLLDDDDCTSAVDGTAMISRTQSRACEKCPALSVDADLIEIGK	13497
QY	368	-----	367	QY	414	-----	413
Db	12418	FNRPEVLGSEAGTPIVIEDEVYEKLSIFCRNHQVTSFVLLAARVAHVHRLTGAEATI	12477	Db	13498	IGFEVASWARQHSQRGGLDVAFHRPEPPRPHSHGVHFRPTEHKGSRSSSLTNRPJLHLQ	13557
QY	368	-----	367	QY	414	-----	413
Db	12478	GTPIANRNRPELDLIGFFVNTQCMRIALEBHDNPLSVRRVRSTAASAFENQDVPPERL	12537	Db	13558	SRLEAKVRERLQSLPPYMIPSRITLLDQMLTNSNGKVDKRLARQARVIPSAASTLD	13617
QY	368	-----	371	QY	414	-----	413
Db	12538	VSALLPGSRASRNPVLQMFVHVSQRNLKQLQLEGLEBTPYTATTRFDVEFHLPEQD	12597	Db	13618	FVAPRTEIEVILCEEFTDLLGVKVGITDNFFELGGHSLIATKLSARLSRRLDAGITVKQV	13677
QY	372	-----	371	QY	414	-----	418
Db	12598	KGLAGNVVFAADLPEAATIRSVVEFHEILLRRGLDQDIALSTWPLVDGLAALNSRLPA	12657	Db	13678	FDQPLADLAASILQOSSRHSRISPSLPEYEGPVEQFAQGRMLWFLQFNFIDALWYLIPFAL	13737
QY	372	-----	377	QY	419	-----	418
Db	12658	VEDIEPDFATEASVVDVFQVWVANDALAVDTSTKLTYAELDQQSDHVAAMLKQKLP	12717	Db	13738	RMRGPLQVDALAAALVALEERHESLRTTFEERDGVGIQVVQPLRTTKDIRIIDVSGMRDD	13797
QY	378	-----	377	QY	419	-----	418
Db	12718	AESIVVVLAPRSSETIVACIGILKANLAYLPMDSNPPEARQALISEIPGEKFFVLGAGV	12777	Db	13798	DAYLEPQKEQOQTPFDLASEPGMRVALLKLGKDDHILSIVMHHIISDGNSTEVLORELQ	13857
QY	378	-----	377	QY	419	-----	419
Db	12778	PIPNKTDVVRMVFISDIVASKTDKSYSPGTRPSASSLAYVIFTSGTGRPKGMVVEHRG	12837	Db	13858	FYLAASKGAPLSQVAPLPIQYRDFAVMQRQEQEVAESQRLDYWKQKQADSSPAELLAD	13917
QY	378	-----	377	QY	420	Y-----	420
Db	12838	VISLVQNASRIPOSRLMAHVSNLAFDASVWEIFPTLLNGGTLFCISYFTVLDSKALSAA	12897	Db	13918	YTRPNVLGEAGSVFVINDSVYKSLVSFCRSRQVTTFTTLAAPRAAAHYRMTGSDDATI	13977
QY	378	-----	384	QY	421	-----	420
Db	12898	FSDHRINITLLPALLKQCLADAPSVLSSLESYIGSDRLDGADATKVXDLVKGKAYNAY	12957	Db	13978	GTPIANRNRPELENLIGCFVNTQCMRITIGDDETFESLVQVVRSTTATAFENQDVPPERI	14037
QY	385	-----	384	QY	421	-----	426
Db	12958	GPTENSVMSTIYTIHETFANGVPIGTSGLPKSKAYIMDQDQQLVPAGVMGELVWAGDGL	13017	Db	14038	VSTLSAGSRDTSRNPVLQLLFAVHSQQGLGRIQLDGVVDEPVLSTVSTFRDLEHAFQEA	14097
QY	385	-----	385	QY	427	-----	430
Db	13018	ARGYDPSLNTGRPIHITIDGKQVQYRGTDRVVRPRDYQIEFFGRLOQOQIKIRGHRIE	13077	Db	14098	DLRNGSVMFATDLFPQETIQGFVAVVEVLQGLEQPOQSIATMPLAEGIAQLRDAGALQ	14157
QY	386	TMFVENKTASDS-----	397	QY	431	-----	435
Db	13078	PAEVEQALLSDSSINDAVVVAQNKEGLEMGVYITTTAAQSVDKKEASNKVQEWEAHFD	13137	Db	14158	MPKSDYPRNASLVDVFOQQAASFTVAVTSTSKLTYAELDRLSQQAASVYLRQQLPAE	14217
QY	398	-----	397	QY	436	-----	437
Db	13138	TAYANIGGIDRDALGDFLSWTSWYDGSILPREMOEWLNDTWRSLLDNQPPGKVLIGT	13197	Db	14218	TWAVLAPRSCETIIAFLAILKANLAYMLDVNTPSARMEAISSVPGRRLLVGVSVHR	14277
QY	398	-----	407	QY	438	-----	437
Db	13198	GTGWLFNLGKVEGLQSYAGLEPSRVTAWVKAETETPFSLAGSARVHVGTAEIDSSIDG	13257	Db	14278	ADINVPNAKTMILISDVTGTDAIGTPBPLVVRPSATSLAYVIFTSGTGKPKGMVVEHRA	14337
QY	408	-----	413	QY	438	-----	437
Db	13258	LRSDLVVINSVAQYFSPREYLAELTANLIRLPGVKRIIFPGDMRTVATNKDFLVARAVHTL	13317	Db	14338	IMRLVKDSNVVTHMPATRMVAHVNTNIAFDVSLPEMCATLLNGGTLVCIDYLTLLDSTMLR	14397
				QY	438	-----	437

Db 14398 ETEREQVRAAIPFPALLROCLVNMPDAICMLBAVYVAGDRFHSRDARATQALAGPRVYN 14457  
QY 438 ---PT---YVN---LTPP---446  
Db 14458 AYGTENAILSTIYNDIKDHPYNGVPIGSAVNSGAYVMDRQQLPFGVMGELVVTGE 14517  
QY 447 ---EH---446  
Db 14518 GVARGYTDASLDTDRFVTTIDGQORAVRTGDRVVRPKGFQIEFFGRLDQAKIRGHR 14577  
QY 447 ---EH---448  
Db 14578 VELGEVEHALLSENSVTDAAVLRTWEEDPQLVAFVTTDHEYRSGSNEEDPYATQAA 14637  
QY 449 ---RAVNLSTGNS---459  
Db 14638 GDMRKELRSLLPYMYVPSRVITILQMLNANGVKDRKDLARRAQMTPTASSGCPVHVAPR 14697  
QY 460 ---459  
Db 14698 NETEAAICDEFETILGVKVIITDNFFELGHSLLATKLAARLSRRMGLRISVKDLFPDPV 14757  
QY 460 ---459  
Db 14758 PVSAGKLEQQQFSGEDESSTVGIVPQLLPAEMSREIIQRDVVPQIENGHSTPLDMYP 14817  
QY 460 ---459  
Db 14818 ATQTOIFFLHDKATGHPATPPLSLDFPETADCRRLASACALVQHFDIFRTVFSRGR 14877  
QY 460 ---459  
Db 14878 FYQVLAHLDPVVEVETIEQELDEVALHEADKQQLRLGRAMLRIALKRPAGKRLV 14937  
QY 460 ---LW---461  
Db 14938 LRMSHLYDGLSLEHIVNALHALYDKHLAQAPKFGLYMHMASRAEGYNFWRSLQGS 14997  
QY 462 ---W---462  
Db 14998 SMTSLKRSVGALEAMTPSAGTWTQTSKIRIPPAALKNGITQATLPTAAVSLLLAKHTKST 15057  
QY 463 ---462  
Db 15058 DVVFGRVSGRODLSINCQDIVGCINEVVRVIDEGDDMGGLLRAIQDQVTSRPHET 15117  
QY 463 ---W 463  
Db 15118 LGLQEVKENC TDW 15130

RESULT 3  
US-09-134-001C-3159  
; Sequence 3159, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3159  
; LENGTH: 10182  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3159

Query Match 21.1%; Score 516; DB 3; Length 10182;  
Best Local Similarity 2.2%; Pred. No. 7.1e-05;  
Matches 225; Conservative 90; Mismatches 129; Indels 9652; Gaps 69;  
QY 2 GRKEMVRDVPKMFVLISISFLLVSEINCKVMS-  
Db 9 GRN-----ICSFLLSKCMYSLSKLTUKTYNFQITSNKKEKTSRI 49  
QY 35 -----34  
Db 50 GVAIALNRRDKLQKPSIRKYAIGTFTSTVIATLVPMTNTHASADELNQNKLIKQLNQT 109  
QY 35 -----34  
Db 110 DDDSNTHSQEIENNKQSSGKTESLSRSTSQANARLSDFKOTNETSQQLPTNVSD 169  
QY 35 -----34  
Db 170 SINQSHSEANMNEPLKVDNSTMQAHSKIVSDSGNASENKHHKLTENVLAESRASKNDK 229  
QY 35 -----34  
Db 230 EKENLQEKDQSOVHPPLDKNALQAFDASYHNYRMI DRDRADATEYQVKVSTFDYVNDL 289  
QY 35 -----KAL-----37  
Db 290 LGNNQNPSEQLVSAVQOLEKALEARTLPQOSTTEKRGRRSTRSVVENRSSRSDYLAR 349  
QY 38 -----YNR-----41  
Db 350 TBYVVS KDDDDSGPPPGTFFHASNRWPYNLPRSRNLRASDVQGNAYITTKRLKDGQOW 409  
QY 42 -----WR 43  
Db 410 DILFNSNHKGHEMYWYFGLPSDQTPGTPVTIINRDSSTSTGCVGFGSGAPLPQWR 469  
QY 44 ---GLVLSKI-----50  
Db 470 SAGAINSSVANDFKHGSATNAYFYDGVNNSDFARGGELYDFREGATQTKYKYGDEFAL 529  
QY 51 -----GKYK-----LDQL 58  
Db 530 LNSEKPDQIRGLDTIYSPKSGDVSYSRISFKTQGTARTLYAAGARSGEYKQATYNNQL 589  
QY 59 KLE-----ILROLETTIS-----71  
Db 590 YVEPYKNYRNVQSNVQVKNRTLHLKRTIRQFDPTLQRTDVPILDGSGSIDSVDPL 649  
QY 72 -----71  
Db 650 SYVKNVGTGVLGIYPSYLPYNOERWQGANMAYQIEELFSQENLQNAARSGRPIQLVG 709  
QY 72 -----TKYNSVK-----78  
Db 710 FDVEDSHHNPETLLPVLNLYVKPELKHTELYHDNEKQDRKFSVSKRAGHGVQVMSGTL 769  
QY 79 -----QPVKN-----LTMN-----87  
Db 770 HNTVSGGILPYQOEIRIKLTSNEPIKDSWSTGYPTNTLQNAVGRTNNAKLNALVG 829  
QY 88 -----TSPF-----91  
Db 830 HDPGNYFITVKFGDVEQFEIRSKPTPRIITANELRGNPNHKEIRVTDIPNDTTAK 889  
QY 92 -----91  
Db 890 IKLVMGTDGDHDPINPYTPENTVVAEYHDNDPSKNGVLTFRSSDYLKDLPLSGEL 949  
QY 92 -----QY-----94  
Db 950 KAIVYNYQVQSNFNSVPFSSDTPPTTINEPAGLVHKKYRGDHEVITLPTVDNTGSGSL 1009

QY 95 ----- 94  
Db 1010 RDVNVNLPQGWTKFTPTINPNNTTEGLKLIGNIPSEAYNTTYHFNITATDNSGNTTNP 1069  
QY 95 ----- 94  
Db 1070 KTFILNVGKLADLPNVLGRDQLQVLTPDSSLSNSERBEVKRISSEANANIRSYLLONN 1129  
QY 95 -ILAG----- 98  
Db 1130 PILAGVNGDVTFYRDGSDVDIDAENVIYEPERKSIIPSENGNTNKEAVITIARGQNT 1189  
QY 99 ----- 98  
Db 1190 IGFNLKRYFSLNSGDLPNRDFTSISAIGSLPSSSEISRLNVGNVYRVNAKNAYHKTQQ 1249  
QY 99 ----- 98  
Db 1250 ELNKLKIVEVNAPTGNRRVYRVSTYLTNDTEINKIKQAFKAANSGLNLNDNDITVSNRF 1309  
QY 99 ----- 98  
Db 1310 DHRNVSVTVTRKGDLLIKEFPSSNLNMNPLRWNIIRDDYTIISWTSKIQGRNTDGLW 1369  
QY 99 ----- 98  
Db 1370 SPDHKSIIYKDATLGRQINTNDVLTLLQATAKNSLNRNSINSEKQLAERGSNGYSKSI 1429  
QY 99 -----PTQ----- 101  
Db 1430 IRDDGSEYLLNSNPICQLDLVEPDNGYGRQVSHSNVIYNEKNSIVNGQVPEANGASA 1489  
QY 102 ----- 101  
Db 1490 FNIQKVVKANAMGIMGVYKRAQLYLPSPKGYIEKLGQNLSTWNVNIVYFVPSDKV 1549  
QY 102 NVSITY----- 107  
Db 1550 NPSITVGNVDHHTVYSGTFKNTINVNDVGLMTVASTDSAITMTNRNNELVGQAPNVT 1609  
QY 108 ----- 107  
Db 1610 NSINKIVKVKATDKSGNESIVSFTVNIKPLNEKYRITTSNNQTPVRISNIQNNANLSIE 1669  
QY 108 ----- 107  
Db 1670 DQNRVKSLSMTKILGRNVYVNESNNDVRSQVSVKUNRSNNATVNVTTTFSDGTTNTIT 1729  
QY 108 ----- 107  
Db 1730 VPKVHVLEVPTRTTVRGQPPTGKGTSPNDFSLRTGPPVDARIIVWVNNQGPDISN 1789  
QY 108 -----LWFPDYSTQLRKPACYVSO-----YNHT----- 131  
Db 1790 QIGRDLTHAEIFFDGETPIRKDTTKLSQSIKPIQIYETTINGRPNSSGDYAPGNFVOA 1849  
QY 132 -----AKTIF-----RPP----- 140  
Db 1850 VNQYWEHMDFRWAGSGTSPSSNAGSFTKTVTVYQNGTENNVNLFVKVKNKFPVDSN 1909  
QY 141 -----PCGRV 145  
Db 1910 SVISKQLNGQQLVRNVPQNAQVTLYQSGNVTIPNTNTTIDSNGIATVTIQTLPTGNI 1969  
QY 146 PSMTCLSEMLNVSKRNDG-----EQCG-N 170  
Db 1970 TAKTSMNNVTYTKQSSGLASNTTDDISVFSNSQVNVNVTAGMQANDGKIIRGTNTN 2029  
QY 171 FTTFNPMFFNVP-----RWN----- 185  
Db 2030 FNDNSFISNIPAHSLTWNBEPSWKNNGITTKTVTLPLPHQGRTRVDPIITIYPTV 2089  
QY 186 -----TKLY----- 189

Db 2090 TAKNPVRDQGRNLNTGTDVNYIIPENNNRLGGTASWKDNQPDKNVAGVQNLIALVNY 2149  
QY 190 ----- 189  
Db 2150 PGISTPLEVPVWVNVNFDFTQPIYKIQVGDFPKGTWAGYKXHLNGBGLPIDGKWFYW 2209  
QY 190 ----- 189  
Db 2210 NQOSTGTTSDQWQSLAYTRTPFVKTGTYDVNPSNWGVWQTSQSAKFIVNAKPNQPTIT 2269  
QY 190 ----- 189  
Db 2270 QSKTGDVTVTPGAVRNILISGTNDYIQASADKIVINKNGNKLTTFVKNDGRWTVETGSP 2329  
QY 190 -----VGPT----- 193  
Db 2330 DINGIGPTNNGTAISLSRLAVRPGDSIEAIAIEGSGETISTATSIEIYVKAPOEQVAT 2389  
QY 194 ----- 193  
Db 2390 HTYDNGTDFILPDNSRNSLNPTERVEINYTEKLNGETQKSFITITKNNGKWTINNKPNY 2449  
QY 194 -KXVDSQTIYFL----- 205  
Db 2450 VEFNQDNGKVFESANTIKPNSQITITPKAGQGNTEINTPTVIPAQAQHTLTINEIVKEQG 2509  
QY 206 ----- 205  
Db 2510 QNVTNDDINNVOVNPKNRVAIKQGNALPTNLAGSTSHIPVVIYSDSGSBEATVTVT 2569  
QY 206 ----- 205  
Db 2570 KVNKTELINARRLDEEISKENKTPSSIRNFQAMNRAQSIQNTAKSDADQVIGTEFATP 2629  
QY 206 -----GLT----- 208  
Db 2630 QOVNSALSKVQAAQNKINEAKALLQNKADNSQLVRAKEQLOQSIQPAASTDGMQDSTRN 2689  
QY 209 ----- 208  
Db 2690 YKNKQAAEQAIQHANSVINNGDATSQINDAKNTVEQAQRDYVEAKSNLRADKSQLQA 2749  
QY 209 ----- 208  
Db 2750 YDTLNRDVLTDKKPASVRRYNEAISNIRKELDTAKADASTLRNTNTPSVEQVRDALNKI 2809  
QY 209 -----ALL----- 211  
Db 2810 NTVQPKVNAQIALLOPKENNELVOAKKRLQDAVNDIPOTQMTQQTINNYNDKQREABR 2869  
QY 212 -LRYAQR----- 217  
Db 2870 ALTSQRVIDNGDATTOEITSEKSKVEQAMQALTNAKSNLRADKNELOQAYNKLINVT 2929  
QY 218 ----- 217  
Db 2930 NGKKPASIRQYETAKARIQNIQINDAKNEAERILGNDNPQVSVQVQALNKIKAIQPKLTEA 2989  
QY 218 -----NCTHSFYLNVAMSR----- 231  
Db 2990 INMLQKENNTLVNAKRNLENVNDTDPHGTQETINNYNAKKREAQNEIQKANMIIN 3049  
QY 232 ----- 231  
Db 3050 NGDATAQDISSEKSKVEQVLQALQNAKNDLRADKRELQYAYNKLQNVNTNGKKPSSIQN 3109  
QY 232 ----- 231  
Db 3110 YKSARRNIENOYNTAKNEAHNVLENTPTVNAVEDALRKINAIQPEVTKAINILOQKEDN 3169  
QY 232 ----- 231

Db	3170	SELVRAKEKLDQAINSQPSLNGCHTQBSINNYTTKREARQNTIASSADTTIINNGDASIEQIT	3229
Qy	232	-NLFRVPKYING-----TKLKNTMRKLAR-----	254
Db	3230	ENKIRVEEATNALNEAKOHLTADTTLSLKTVEVKLSRRGDTNNKKPSSVSAYNNTHISLQS	3289
Qy	255	-----KOAPVKEQ-----	254
Db	3290	EITQTENRANTIINKPIRSVEEVNNAHEVNQNLQRLTDTINLLQPLANKESLKEARNRL	3349
Qy	255	-----KOAPVKEQ-----	262
Db	3350	ESKINETVQDGMTQOSVENYKOAKIKQAQESSIAQTLINNGDASDQEVSTEISKLNOKL	3409
Qy	263	-----FEKK-----	266
Db	3410	SELTNSINHLTVNKBLETAKNQLQANIDQKPSDGMTQOSVQSERKLEAKDKINSIN	3469
Qy	267	-----	266
Db	3470	NVLANNPDVNAIRTNKVETEQINNELTOAKOGLTVDQPLINAKTALQOSLIDNQPFSTGM	3529
Qy	267	-----AKK-----	269
Db	3530	TEATIQYNAKQKAEQVIONANKIENAPQSVQVSEKSKVEQALSELNNAKSALRAD	3589
Qy	270	-----YTT-----	269
Db	3590	KQELQAYNQLIPTDLNNKPASITAYNQRQOQFSNELNSTKTNDRLKEQNPVADV	3649
Qy	270	-----TOSTTTPYFS-----	279
Db	3650	NNALNKVREVQKLEARNALLQNKEDNSALVRAKEQLQAVDQVPSTEGMTQOTKDDYNS	3709
Qy	280	-----YTT-----	282
Db	3710	KQAAQOEISKAQVINDGDATTQOISNAKTNVERALEBALNAKTLGLRADKEELQONAYNQ	3769
Qy	283	-----SAALN-----	287
Db	3770	LTQNDITSGKTSPASIRKYNEAKSRITQDIDSAKNEANSILTNDNPQSVQVTAALNKIKAV	3829
Qy	288	-----	287
Db	3830	QPELDKAIAMLNKNNNALVQAKQLOQIVNEVDPTQMTTDTANNYKSKREADEBIQ	3889
Qy	288	-----	287
Db	3890	KAQIINNGDATEQQITNETNRVNOAINAINKAKNDLRADKSQLENAYNQLIONVDTNKG	3949
Qy	288	-----	287
Db	3950	KPASIOQYQARQAIETQYNNNAKSEAHQILENSPVSNEVAQLOKVEAVOLKUNDIHI	4009
Qy	288	-----	287
Db	4010	LQWKENNSALVAKNQLQOSVNDQPLTTGMTQDSINNYEAKNEAQSAIRNAEAVINNGD	4069
Qy	288	-----	287
Db	4070	ATAKQISDEKSKVEQALAHLNDAKQOLTADTTTELQTAVQQLNRRGDTNNKKPRISINAYNK	4129
Qy	288	-----	287
Db	4130	ATQSLETQITSAKDNANAVIQPIRTVQEVNNALQOVNQLNQLTEALINQLQPLSNDAL	4189
Qy	288	-----	287
Db	4190	KAARLNLNKINQVTDGMTQOSIEAYQNAKRVQAKNESNTALALINNGDADEQQIITET	4249
Qy	288	-----VTNNVTYSI-----	296
Db	4250	DRVNOOTNLTOAINGLTYNKKEPLETAKTALONNIDOVFSDGMTQOSVANNOKLOIAK	4309
Qy	297	-----	296
Db	4310	NEINTINNVLANNPDVNAIKTNKAEARISNDLTQAKXNLQVDTQPLEKIKRQLQDEIDQ	4369
Qy	297	-----	296
Db	4370	GTNTDGMTQOSVDNYNDSLSAAITEKGKVNKLKRNPFTVEQKESVANAQOVIQDLQNAK	4429
Qy	297	-----	296
Db	4430	TSLSVPDKTQLEAKNRLSNSINQOTDGTQDSLNYYNDKLAKARQNLKESIKVLGGQP	4489
Qy	297	-----	296
Db	4490	TVAEIRQNTDEANAHKQALDARSQTLNREPIYINHINNESHNLNAQDNFKAQVNSAPN	4549
Qy	297	-----	296
Db	4550	HNTLETIRKVKADTLNQSMTALSSEIADYENOKQENYLDASNNKRQDYDNVNAAKGILN	4609
Qy	297	-----TTAARRVSTST	307
Db	4610	QTSQPTMSADVIDQKAEVDKRTKTALDGNQRLEVAKQQAALNHLNTLNDLNDARQOTLTDT	4669
Qy	308	IAYRP-----	312
Db	4670	INHSPWINSVQAKKANTVNTAMTQLKQTIANYDYDELHDGNYINADKKDAYNNAVNN	4729
Qy	313	-----	312
Db	4730	AKQLINQSDANQAQDPABEINKVTQRVNTTKNDLNGNDKLAEAKRDANTTTDGLTYLNEA	4789
Qy	313	-----	312
Db	4790	QRNKAKEVGRKASTKTNITTSQLODYNQLNIAQALRNSVNDVNVKANSYINEDNGPKE	4849
Qy	313	-----	312
Db	4850	AYNQAVTHAQTLINAQSNPEMSRDVVNQKTQAVNTAHQNLHGQKLEQAQSSANTEIGNL	4909
Qy	313	-----DSSF--MKSMTAQ-----	324
Db	4910	PNLTNTQAKKEKELVNSKQTRTEVQEQNLQAKSLDSSMGTLSLVAKQPTVQKTSVYINE	4969
Qy	325	-----	324
Db	4970	DOPEQAYSNDISITMGQTIINTKADPVLDKTLVDNAISNISTKENALHGEQKLTAKTEAI	5029
Qy	325	-----	324
Db	5030	NALNTLADLINTPOKEAIKTAINTAHTRTDVTAEQSKANQINSAMHTLRQNISDNESVTNE	5089
Qy	325	-----	324
Db	5090	SNYINAEPEKOHAFTEALNNAKEIVNEQOATLDANSINQAKAAILTTKNALDGBEQLRRA	5149
Qy	325	-----	324
Db	5150	KENADQEIINTLNLTAORNSEKGLINSSQTRTEVASQAKAKELNKVMEQLNHLINGKN	5209
Qy	325	-----	324
Db	5210	QMINSSKFINEDANQQOQAYSNAIASAEALKNKSNQPELDKVTIEQAINNINSAINNLNGE	5269
Qy	325	-----LRD	327
Db	5270	AKLTRAKEDAVASINNLGLTNEQKTENQAVNQATRDQVANKLRDAEALDQSMQTLRD	5329
Qy	328	LA---TWYTTILRY-----RQNP-----	342
Db	5330	LVNNQNAIHSTSNYFNEDSTQKNTVDNAIDNGSTYITQCHNPENLKFSTIDQITSRINTAK	5389

QY 343 ----- 342  
Db 5390 NDHLGVEKLQDKGTANQEIGQLGYLNDPKQSGEISLVNGSNTRSEVEEHLNEAKSLNNA 5449  
QY 343 ----- 342  
Db 5450 MKQLRDKVAEKTNVKSSDYINDSTEHQRYDQALQAEANIINEIGNPTLNKSEIBOKLQ 5509  
QY 343 ----- 342  
Db 5510 QLTDQNALQSGHLLBEAKNAITGINKLTALNDAQROKAIENVQAQOTIPAVNQQLTLD 5569  
QY 343 ----- 342  
Db 5570 REINTAQAALRDVKGQONNVHQSNYFNEDEQPKHNYDNSVQAGQTIIDKLQDPIMNKNE 5629  
QY 343 ----- 342  
Db 5630 IEQAINQINTTQALSGENKLHTDQESTNRQIEGLSSLTAQINAEKDLVNAQAKTRDVA 5689  
QY 343 ----- 342  
Db 5690 QKLAAAKEINSAMSLNRDGIQNKEDIKRSSAYINADPTKVYTDQALQNAENIINATPNV 5749  
QY 343 ----- 342  
Db 5750 ELNKATIEQALSRVQAQDLDGVQOLANAKQATQTVNGLNSLNDGQKRELNLLINSAN 5809  
QY 343 ----- 342  
Db 5810 TRTKVOEELNKATELNAHEALRNSVQNVQVQKSSNYNVEDQPEQHNYDNVNEAQATI 5869  
QY 343 ----- 342  
Db 5870 NNNAPVLDKLAIERLTQTVNTTKDALHGAQKLTQDQQAETGIRGLTSLNEPKNAEVA 5929  
QY 343 ----- 342  
Db 5930 KVTAATTRDEVRNIRQBEATLTDAMLGLRSIKDKNDTKNSKYINEDHDQQQAYDNVAVN 5989  
QY 343 ----- 342  
Db 5990 NAQOVIDEQATLSSTINOLANAVTAQAKSNLHGDTKLQHDKDSAKQTTIAQQLNSAQK 6049  
QY 343 ----- 348  
FcPSR-----  
Db 6050 HMEDSLIDNESTETQVQHDLTEAQAQDGLMGALKESI KDYTNIVSNGNYINAEPSKKQAY 6109  
QY 349 ----- 360  
NRTAVSEFMKVT-----  
Db 6110 DAAVQNAQNIINGTNOPTINKGNVTATQTVKNTKDALDGHRLBEAKNNANQTI RNLSN 6169  
QY 361 ----- 360  
Db 6170 LNNAQDAEKNLVNSASTLEQVQONLQTAQQLDNAMGELRQSI AKKDQVKADSKYLNEDP 6229  
QY 361 ----- 360  
Db 6230 QIKQNYDDAVQVETIINETONPELLKANIDQATQSVQNAEQALHGAEBKLNQDKQTSSTE 6289  
QY 361 ----- 360  
Db 6290 LDGLTDLTAQREKLEQINTSNSRDDIKQIEQAKALNDAMKLEQVAQKDGVHANGD 6349  
QY 361 ----- 360  
Db 6350 YTNEDSAQDAYNNALKQAEIDIINSSNPNLNAQDITNALNNIKQAQDNLHGAQKLQDDK 6409  
QY 361 ----- 360  
Db 6410 NTTNQAIGNLHNLNQPKDALIQAINGATSROQVAEBKLEAEALDEAMKQLEQVQVNDQDQ 6469  
QY 361 ----- 360

Db 6470 ISNSSPFINESDKQKYNDKIQAAKEIINQTSNPTLDKQKIADTILQNIKDAVNNLHGDQ 6529  
QY 361 ----- 360  
Db 6530 KLAQSKQDANNQLNLHLDLTERQKQNHFKPLINNADTRDEVNKQLEIAKQLNGDMSTLHKV 6589  
QY 361 ----- 360  
Db 6590 INDKQIQHLSNVINADNKQNYDNAIKEAEDLIHNPPTLDHKKALQDLNLNKIDQAHNE 6649  
QY 361 ----- 363  
HVL-----  
Db 6650 LNGESFKQALDNALMDIDSLSNLVNPQRQTVKDNINHVTTLESLAQELQAKELNDAMK 6709  
QY 364 ----- 371  
IRNETPYT-----  
Db 6710 AMRDSIMNQEQIRKNSNYTNEDLAQONAYNHAVDKINHIIGEDNATMDPOI IKAQTODIN 6769  
QY 372 ----- 371  
Db 6770 TAINGLNGDQKLQDAKTDKQITNPTGLTEPQKQALENIINQOTSRAVNAKQLSHAKFL 6829  
QY 372 ----- 371  
Db 6830 NGKMBELKVAVAKASLVQRNSVYNEDVSEKEAYEQIAKQOEIINSENNPTISSTDINR 6889  
QY 372 ----- 374  
IYG-----  
Db 6890 TIOEINDAEONLHGDKNLQAOEIAKNEIQNLDSLNSAQITKLIQIDIGRTTKPAVTQKL 6949  
QY 375 ----- 377  
TLD  
Db 6950 EEAKAINQAWQKLQKSIADKDATLSSNYLNEDSEKKLAYDNAVSAEQILNQLNDPTMD 7009  
QY 378 MSSL----- 381  
Db 7010 ISNIOAITOKVIAKDSLHGANLQONQADSNLIINQSTNLNDKQKQALNDLINHAQTKQ 7069  
QY 382 ----- 391  
YNETMFVEN-----  
Db 7070 QVABIIAQANKLNEMGTCLKTIVEQSNVHQSKYINEDPQVQNIYVNDSIQKGREILNGT 7129  
QY 392 ----- 391  
Db 7130 TDDVLNNNKIADAIQNIHLTKNDLHGDQKLOKAQQDATNELNYLTNLNNSQRQSEHDEIN 7189  
QY 392 ----- 391  
Db 7190 SAPSRTEVSNLNLHAKALNEAMROLENEVALENSVKLSDFINEDBAQNEYSNALQKAK 7249  
QY 392 ----- 391  
Db 7250 DIINGVPSSTLDKATIEDALLELQNARESLHGEQKLOEAKNOQAVAEIDNLQALNPGQVLA 7309  
QY 392 ----- 391  
Db 7310 EKTLYNQASTKEVQEQALQAKELNEAMKALKTEINKKEQIKADSRVYNADSLQANYNS 7369  
QY 392 ----- 400  
KTA-----SDSKNT-----  
Db 7370 ALNYGSIITATTQPPPELNKDVINRATQTIKTAENNLNGSKLAEAKSDGNSQSTIEHLQGLT 7429  
QY 401 ----- 400  
Db 7430 QSQKQKHDLINQAQTKQVDDIVNNSKOLDNSMNQLOQIIVNNDNTVKQNSDFINEDSSQ 7489  
QY 401 ----- 400  
Db 7490 QDAYNHIAQAKDLITAHPTINDKQIQDOAIENIKQALNDLHGSNKLSEDKKEASFQON 7549  
QY 401 ----- 400

Db 7550 LNSLTNGQDQTLNHIHIFSPATRSQVGEKIASAKQLANTMTKALRDSIADNNEILQSSKYFN 7609  
QY 401 -----TP----- 402  
Db 7610 EDSEQNAYNAVNKAKNIINDQPTPVMANDEIQSVLNEVKQTKDNLHGDKLANDKTDA 7669  
QY 403 ----- 402  
Db 7670 QATNALANYLNAQRNLETKVQNSRPRPEVKVQLANQLNDAMKCLDDALTGNDAIKQ 7729  
QY 403 -----TSPM----- 407  
Db 7730 TSNYINEDTSQVNFDEYTRGKNIIVAEQTNPNMSPNTINTIADKITEAKNDLHGQKUK 7789  
QY 408 ----- 407  
Db 7790 QAQQQSINTINQMTGLNQAQKEQLNQEIOQTQTRSEVHVQVINKAQAALNDSMTLRQSITD 7849  
QY 408 ----- 407  
Db 7850 EHEVKQTSNYINETVGNQTAYNNAVDKVKIINQTSNPTMNPLEVERATSNVKISKDALH 7909  
QY 408 ----- 407  
Db 7910 GERELNDKNKSTFAVNLHDLNLAQAQKEALTHEIEQATIVSQVNNIYNKAKALNNDMMKL 7969  
QY 408 ----- 407  
Db 7970 KDIVAQODNVRQSNYINEDSTPQNMVNDTINHAQSIIDQVANPTMSHDEIENAINNIKH 8029  
QY 408 ----- 407  
Db 8030 AINALDGEHKLQAQKENANLLINSLNDLAPQDAINRLVNEAQTRKVAEQLSQAALN 8089  
QY 408 ----- 407  
Db 8090 DAMKHLRNSIQNSSVRQESKYINASDAKKEQVNHAVREVENIINEQHPTLDKEIHKQLT 8149  
QY 408 ----- 407  
Db 8150 DGVQANNDLNGVELLDADQNAHQSIPTLHLNLAQAQNALNEKINNVAITRTEVAALIGQ 8209  
QY 408 ----- 407  
Db 8210 AKLLDHAMENLEBSIKDKEQVQSSNYINEDSDVQETYNADVHVTIILNQTVNPTLSIE 8269  
QY 408 ----- 407  
Db 8270 DIEHAINEVNAQKQLRGKQKLYQTTIDLADKELSKLDDLTSQQSSSISNQIYTAKTRETV 8329  
QY 408 ----- 407  
Db 8330 AQAIKAKSLNHAMKALNKVKNADKVLDSRFINEDQPEKKAYQQAINHVDISIHRQTN 8389  
QY 408 ----- 407  
Db 8390 PEMDPTVINSITHELETAQNHLGDKLAHAQDAANVINGLIHLNVAQREVMTNTNA 8449  
QY 408 ----- 407  
Db 8450 TTRKRVAKNLDNAALDKAMETLQQVVAHKQNILNDSKYLNEDESKYQQQYDRVIADEQL 8509  
QY 408 ----- 407  
Db 8510 LNQTTNPTLEPYKVIDIVKQNVLANEKILFGAEKLSYDKSNANDEIKHMYLNNAQKQSIK 8569  
QY 408 ----- 407  
Db 8570 DMI SHAALRTEVKQLLQAQKILDEAMKSLDKTQVWITDTPLNTEASEDKKEKQVDTV 8629  
QY 408 ----- 407  
Db 8630 SHAQAIIIDKINGSNVSLDQVQRALEQTOASENLGDGQRVVEAKVHANQTIQDLTHLSL 8689

QY 408 ----- 407  
Db 8690 QOQTAKESYKNAKLEBIATVSNNQAALNKMVGKLEQFINHADSVENSQNYROADDDKII 8749  
QY 408 -----GFOR----- 411  
Db 8750 AYDEALEHGDIOKTNATQNETKQALQOLIYAETSINGFERLNHARPRALEYIKSLEKIN 8809  
QY 412 ----- 411  
Db 8810 NAKSALEDKVTQSHDILLELEHIVNEGTVNLDMGELANAIVNNVAPTASINYINADNL 8869  
QY 412 -----TFIDPLW-----DYL 421  
Db 8870 RKDNFTQAINNARDALNKTQGNLDFNAIDTTFKDDIFKTKDALNGIERLTAASKAKAEKII 8929  
QY 422 DSIILFL-----DETRNF-----SLR-----SPT 439  
Db 8930 DSLKFINKAQFTHANDEIMNTNSIAQLSRIVNOAFDLNDAMKSLRDELNNQAPFVQASSN 8989  
QY 440 YVN----- 442  
Db 8990 YINSDLEDKQFDFHALSNARKVLAKENGKNLDEKQIOGLKQVIEDTKDALNGIQLRSKAK 9049  
QY 443 ----- 442  
Db 9050 AKAIQVVSLSYINDAQRHIAENNIHNSDDLSSLANLTLKASDLNADNMDLRDTIESNST 9109  
QY 443 ----- 442  
Db 9110 SVPNSVNYINADKNLQIEPDEALQOASATSSKTSNPATIEEVLGUSQAIYDTGNALNGB 9169  
QY 443 ----- 442  
Db 9170 QRLATEKSKDLKLIKGLKDLNKAQLEDVTNKVSANTLTLSQLTQSTLELNDKMKLLRD 9229  
QY 443 -----LTPP----- 446  
Db 9230 KLKTLVNPVKASLYRNADYNLKRQFNKALKEAGVLNKNSGTNVNINDIQHLLTQIDNA 9289  
QY 447 -----EHR----- 450  
Db 9290 KDQLNGERLKEHQKSEVFIKELDILNNAOKAAIINOIRASKDIKIINOIVDNAIELN 9349  
QY 451 -----AVNL----- 454  
Db 9350 DAMQGLKEHVAQITATTKDNIEVLNADHDKLYDYVAINLANNVLDKENGTKNDANIIG 9409  
QY 455 -----STNS----- 459  
Db 9410 MIQNMDARALLNGIERLKDAQTKAHNDIKDTLKRQLEBIEHANATSNKSAQAKQMVNEE 9469  
QY 460 ----- 459  
Db 9470 ARKALSNINDATSNDLVNOAKDEGOSAIIEHHADELFPKAKLDANQMDOKVEDINHLSQ 9529  
QY 460 ----- 459  
Db 9530 NPNLSNEEKKNKLSIQINKLVNGIKNEIQQAINKQOQIENATTKLDEVIETTKKLIIAKAEA 9589  
QY 460 ----- 459  
Db 9590 KQMIKELSQKRDAINNNNTDLTPSKAHALADIDKTEKDALQHIENSNSIDINNNEHA 9649  
QY 460 -----LW----- 461  
Db 9650 FNTLHIIIIWTDQQLVFEVPELSLQNALVTSVVVVRDETISLESIKKKMTLTDLKV 9709  
QY 462 ----- 461  
Db 9710 NIVSLPNTDKVADHLTAKVKVILADGSYTVNVVPKVVKEKELQIAKDAIKTIDVLVKQK 9769

```
Qy 462 ----- 461
Db 9770 IKDIDNNELTSTOREDAKAEIERLKQKQIDKVNHSKSIKDIETVKRTDPEIDQDPKR 9829
Qy 462 ----- 461
Db 9830 FTLNKAKKDITDVNTQIONGPKEIETIKGLTSNEKTQPKQLTALQKEFLEKVEHAHL 9889
Qy 462 ----- 461
Db 9890 VELNQLQOEFNNRYKHILNQAHLLGEXHIAEHKLGIVVNNKTOQILNNQSASYFIKWAL 9949
Qy 462 ----- 461
Db 9950 DRIKIQIETWNSIRGAHTVQDVHKKALLQIEQILKVNVSIIINQSFNDSLHNFYLHSKF 10009
Qy 462 ----- 461
Db 10010 DARIREKDVANHIVQTFEVLKGTGVEPKINKETQPKLHKNDNDSLFKHLVDNFGK 10069
Qy 462 -----WWL 464
Db 10070 TVGVITLTGLSSFWL 10085

RESULT 4
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-6

Query Match 19.7%; Score 483; DB 3; Length 11877;
Best Local Similarity 2.0%; Pred. No. 0.0013;
Matches 229; Conservative 77; Mismatches 149; Indels 11020; Gaps 74;

Qy 2 GRKEMVRDVPKM----- 14
Db 217 GLAEILRRHIPKRTYRSLADRHGIPDGGLPKDILTLQWIALFOASQPSYAPGAPGTRMPG 276
Qy 15 ----- 14
Db 277 QGGAGGRDYDSETSRAAVGSSRRYGTTRGGEPCAPRAQVROTQKRGARGSSYGRRTGR 336
Qy 15 ----- 14
Db 337 MSSAGITRTGARTPTVTRGNAAWDTGEVRVRRLGPPAGPDHABHSFSAFTGTVRAELIR 396
Qy 15 -----FVLIS----- 19
Db 397 GEMSTVSKSESEBFVSVNDAGSAHGTABPVAVVGISCRVPGARDPREFWELLAAGQAV 456
Qy 20 ----- 19
Db 457 TDVPADRWNAGDYPDRSAPGRNSRWGGFTIEDVDRFDAFFGISPREAAEMDPQORLA 516
Qy 20 ----- 19
Db 517 LEIGWEALERAGIDPSSLTGTGTGTFAGAIWDDYATLKHRQGGAAITPHTVTGLHRGIITA 576
```

```
Qy 20 -----ISFLV----- 25
Db 577 NRLSYTLGLRPSMVVDSCQSSSLVAHVHLACESLRGESELALAGGVSLNLVPSDSIGAS 636
Qy 26 -----SFINCKVMSKAL----- 37
Db 637 KFGGLSPDGRAVTFDARANGYVRGEGGFVWLKRLSRAVADGDPVLAVIRGSVAVNGGAA 696
Qy 38 ----- 37
Db 697 QGNTTPDAQOQBAVLREAHERAGTAPADVRYVELHGTGTPVGDPIEAAALGAALGTGPA 756
Qy 38 -----YNRP----- 41
Db 757 GQPLLGVSVKTNIGHLEGAAGTAGLIKAVLAVRGRALPASLNYETNPAPFPEELNLRVN 816
Qy 42 ----- 41
Db 817 TEYLPWEPEHDAQRMVGVSSFGMGGTNAHVLEAPGCGRGASVVESTVGGSAVGGVV 876
Qy 42 ----- 41
Db 877 PWVSAKSAALDAQIERLAASFASRDTGVDAGAVDAGAVARVLAGRAQFEHR 936
Qy 42 -----WRGL----- 45
Db 937 AVVVGSGPDDLAALAAPESGLVGVASGVGRVAFVPPGQGTQWAGMGAELLDSSAVFAA 996
Qy 46 ----- 45
Db 997 MAECAAALSPYVDWSLEAVVRQAPGPTLERVDVQPTFAVMVSLARVMQHHGVTPOAV 1056
Qy 46 -----VLSKIGY 53
Db 1057 VGHSGEIAAAVAGALSDDAARVVTLRSKSIAAHLAKGKGMLSALSEDVLERLAGF 1116
Qy 54 -----KLDQKLEIL-----RQLE----- 67
Db 1117 DGLSVAAVNGPTATVVSGDPVQIEELARACEADGVRARVIPVDYASHSQVEIIESELAE 1176
Qy 68 -----TTISTK-----Y 74
Db 1177 VLAGLSPOAPRVPPFFSTLEGAMITEPVLDDGGYVYRNLHRVGFAPAVETLATDEGTFHV 1236
Qy 75 NVSKOPVKNLTWN----- 87
Db 1237 EVSAHPV--LTWALPGTVTGLATLRDNGGQDRLVASLAEAWANGLAVDWSPLPSATGH 1294
Qy 88 ----- 87
Db 1295 HSDLPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTAEABEPAELDRDQLRVLDKVRA 1354
Qy 88 ----- 87
Db 1355 QTAQVLGYATGQIEVDRTFREAGCTSLTGLDRLNRINAAGFVRMAPSMIFDFTPEALA 1414
Qy 88 ----- 87
Db 1415 EQLLLVHGEAAAANPAGAPAPVAAAGAVDEPVAIVGMACRLPGGVASPEDLWRLVAGGG 1474
Qy 88 -----TEFPQ----- 92
Db 1475 DAISEFPQDRGWDEGLYHPDPEHPGTSYVRQGGFIENVAGFDAFFGISPREALAMPQ 1534
Qy 93 -----YYILAG----- 98
Db 1535 QRLLETSWAEVADAGIDPTSLRGRQGVGTGAMTHEYGPSLRDGGEGDGLYLLTGNTAS 1594
Qy 99 ----- 98
Db 1595 VMSGRVSYTLGLEGPALTVDTACSSSLVALHLVAQLRKGEVDMALAGVAVMPTGPMFV 1654
```



QY 99 ----- 98  
Db 1655 EFSRQRLAGDRSKAPASADGTSNSEGVLVLLVERLSDARNRNGHVLAVRGSAHQD 1714  
QY 99 -----PI----- 100  
Db 1715 GASNGLTAPNGSQORVIRBALADARLTTSDDVDVVEAHGTGRLGDPFIEAQALATYGG 1774  
QY 101 ----- 100  
Db 1775 RDDEQPLRLGSLKSNIGHTQAAAGVSVIKWVQAMRHGLLPKTLHVDEPSDQIDWSAGAV 1834  
QY 101 ----- 100  
Db 1835 ELLTEAVDWPEKODGGLRAAVSSFGISGTNAHVLEAPVVPVVEGASVVEPSVGSAGVG 1894  
QY 101 ----- 100  
Db 1895 GVTTPWVSAKSAALDAQIERLAASFASRORTDDADAGAVDAGAVAHVLADGRAQFEHRAV 1954  
QY 101 ----- 100  
Db 1955 ALGAGADDLVQALADPDGLIRGTASGVRVAFVFPQGTQWAGMGAELLDSSAVFAAAMA 2014  
QY 101 -----QYYSIT----- 106  
Db 2015 ECEAALSPYVDMNSLEAVVRQAPGAPTLEVVVQVPTFVPMVSLARVWQHGVTPQAVVG 2074  
QY 107 ----- 106  
Db 2075 HSQGEIAAYVAGALPLDDAARVVTLRSKSIAAHLACKGMSLALNEDAVLERLSDPDG 2134  
QY 107 ----- 106  
Db 2135 LSVAAVNGPTATVSGDPVQIBELAQACKADGPRARIIPVDYASHRSQVEIIESELAQVL 2194  
QY 107 ----- 106  
Db 2195 AGLSPOAPRPPFTLEGTWITPEVLDTGYWYRNLRRHVGFAPAIETLAVDEGTFHFVEV 2254  
QY 107 ----- 106  
Db 2255 SAHPVLTWLTPEVTGLTLRREQQGERLVTSLAEAWNGLPVAVTSLLPATASRPLP 2314  
QY 107 ----- 106  
Db 2315 TVAFQERYWLENTPAALATGDDWRYRIDWKRLPAEAGSERTGLSGRWLAVTPEDHSAQA 2374  
QY 107 ----- 106  
Db 2375 AAVLTALVDAGAKVEVLTAGDDDDREALAARLTALTGDTGWVSLDGLVPQVAWQA 2434  
QY 107 ----- 106  
Db 2435 LGDAGIKAPLNSVTQASVSGRLDTPADPDRAMLWGLGRVVALEHPERWAGLVDLPQAPD 2494  
QY 107 ----- 106  
Db 2495 AAALAHVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRRPTRDWQPHGTVLITGGTGA 2554  
QY 107 ----- 106  
Db 2555 LGSHAARMMAHGAHELLVRSRGEQAPGATQLTABLTASGARVTIAACDVADPHAMRTL 2614  
QY 107 ----- 106  
Db 2615 LQAIAPETPLTAVVHTAGALDDGIVDTLTABQVRRRAHRAKAVGASVLDLTDRLDLDAFV 2674  
QY 107 -----YL----- 108  
Db 2675 LFSSVSSTLGIPOGNYAPHNAVLDALAAARRATGRSAVSVANGPMDGGMAAGDQVAER 2734  
QY 109 -----WDFY----- 113

Db 2735 LRNHGVPQMDPELALAALESALGRDETAITVADIDMDRFTLAYSSGRPOPLVEBELPEVRR 2794  
QY 114 ----- 113  
Db 2795 IIDARDSATSGOGSSAQGANPLAERLAAAAFCERTEILLGLVRAQAAAALVLRMRSPEDVA 2854  
QY 114 -----STQLRKPKYVY----- 125  
Db 2855 ADRAFMDIGFDSLAGVELNRLTRATGLQPATLVFDPHTPLALVSLLRSEFLGDEBTAD 2914  
QY 126 ----- 125  
Db 2915 ARRSALPATVAGAGAGAGTADDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGSGITP 2974  
QY 126 ----- 125  
Db 2975 FPTDRGMDLGLYDADPDALGRAYVREGGFLHDAAEFDEFFGVSPREALAMPDQORMLL 3034  
QY 126 ----- 125  
Db 3035 TTSWEAFERAGIEPASLRGSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGR 3094  
QY 126 ----- 125  
Db 3095 IAYTFLEGPATVDTACSSSLTALHLAVLRALRSRECTMALAGGVAMMATPHMFVFSRQ 3154  
QY 126 ----- 125  
Db 3155 RALAPGRSKAFSADADGFGAAEGVCLLLVERLSDARRNGHPVLAVVGTAVNQDGASNG 3214  
QY 126 ----- 125  
Db 3215 LTAPNGSQORVIRQALADARLAPGIDAVETHGTGTSIGDPIEAOGLQATYKERPAAER 3274  
QY 126 -----SQYNT----- 131  
Db 3275 PLAIGSVKSNIGHTQAAAAGAGIIRKVLAMRHGTLPTKLHADEPSPHVDWANSGLALVTE 3334  
QY 132 ----- 131  
Db 3335 PIDWAGTGPRAAVSSFGISGTNAHVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGT 3394  
QY 132 ----- 131  
Db 3395 SEVAEGSEASEAPAPGSGREASLPGLHPWLAKDBEQSLRQAAALHAWLSEPAADLSA 3454  
QY 132 ----- 131  
Db 3455 DGPALRDVGYTLATSRTAFAHRAAVTAADRGFLDGLATLAQGGTSAHVHLDLTARDGTT 3514  
QY 132 ----- 131  
Db 3515 AFLPTQSSQRPAGRELYDRHPVAFARALDEICAHLDGHLELPLLDVMPAAEGSAEAALL 3574  
QY 132 ----- 131  
Db 3575 DETRYTQCALFALEVALFRLVESWGMPPAALLGHSVGEIAAAHAGVAFSLADAARLVAAR 3634  
QY 132 ----- 131  
Db 3635 GRLMQELPAGGAMLAQAAEDEIRVWLEETEYAGRLDVAANGPEAAVLSGDADAAREA 3694  
QY 132 -----AKTITPRPP----- 140  
Db 3695 EAYNSGLGRRTALRVSHAFSAHMDGMDGDFRAVLETVETVEFRPRPSLTVVSVNTGLAAGPD 3754  
QY 141 ----- 140  
Db 3755 DLCDPEYVVRHVGTFRFLDGVRVLRDLGVRTCLELGPDLVTAMAADGLADTADTADSAAG 3814  
QY 141 ----- 140

```
Db 3815 SPVGPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAG 3874
Qy 141 ----- 140
Db 3875 SGAHRVLDPTYSPFRDRYLDAPAADTAVDTAGLGTADHPILLGAVVSLPDRDGLLLTG 3934
Qy 141 ----- 140
Db 3935 RLSLRTHPLADHAVLGSVLLPGAAVLAHAHAESAGLRDVRLELTLEPLVLPHEGGVE 3994
Qy 141 -----PCGR----- 144
Db 3995 LRVTVGAPGPGESAGDGARPVSLHSLRADAPAGTAMSWCHATGLLATDRPELPVAPDR 4054
Qy 145 ----- 144
Db 4055 AAMWPPQGAEBVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATNTATAP 4114
Qy 145 ----- 144
Db 4115 ATANGGSAAPYGIHPALLDASLHAIAVGLVDEPELVRVPFWHSVTVHAAGAAAR 4174
Qy 145 ----- 144
Db 4175 VRLASAGTDAVSLSLTDGGRPLVSVVERLTLPVTADQAAASRVGGLMHRVAMRPYALAS 4234
Qy 145 -----VPSMTCLSE----- 153
Db 4235 SGRQDPHATSYGPTAVLGKDELKVAALAESAGVEGLYPLDALALSQDVAAGAPRTVLA 4294
Qy 154 ----- 153
Db 4295 PLPAGPADGAGVGRTVARTLLEQLAWLADEHLAQLLTVRGAVRDPGSGADGGE 4354
Qy 154 ----- 153
Db 4355 DLSHAAWGLVRTAQTENPGRFGLLDLADDASSYRTLPSVLSDAGLRDEPQLALHDGTTIR 4414
Qy 154 -----MLNVSKR----- 160
Db 4415 LARLASVRPETGTAAPALAPEGTVLLTGGTGGGLGVARHVWGVRRLLLVSRGTD 4474
Qy 161 ----- 160
Db 4475 PGADLVHELEALGADVSAACDVADREALTAVLDAIPAHEPLTAVVHTAGVLSGDTLPS 4534
Qy 161 ----- 160
Db 4535 MTTEDVEHVLRPKVDAAPFLDDELSTPAYDLAAPVMPSSAAAVFGAGQGAYAAANATLD 4594
Qy 161 ----- 160
Db 4595 ALAWRRRAAGLPALSLGWLWETSMTGELQADLRRMSRAGIGGISDAEGIALLLDAL 4654
Qy 161 -----ND----- 162
Db 4655 RDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRRPSSAASATTAGTAGT 4714
Qy 163 ----- 162
Db 4715 PGTDGAETAATVTLADRAATVDGPARQRLILLEFVVVGEVAEVLGHARGHRI DAERGFLDL 4774
Qy 163 ----- 162
Db 4775 GFDSLTAVELNRNLNSAGGLALPATLVFDHPSPALASHLDAELPRGASDQDAGNRNGN 4834
Qy 163 -----TGQGCN 170
Db 4835 ENGTTASRTAETDALLAQLTRLEGALVLTGLSDAPGSEVLEHLRSLSMVTGTGTGT 4894
Qy 171 FT----- 172
Db 4895 ASGAPDGAGSGBDRPWAAGDGAGGSGEDGAGVPDFPMNASAEELFGLLDQDPSTDMSTVN 4954
```

---

```
Qy 173 ----- 172
Db 4955 EEKYL DYLRATADLHEARGRLRELEAKAGEPVAIVGMACRLPGGVASPEDLURLVAGGE 5014
Qy 173 -----TFNPFNFV-PR----- 183
Db 5015 DAISEFPQDRGMDVEGLYDPNPEATCKSYAREAGFLYEAGFDADFFGISPREALAMDPO 5074
Qy 184 -----WN-----TKL-----YVGP----- 192
Db 5075 QRLLEASWEAFEHAGIPAAATARGTSVGVTGMVMDHYATRLTDVPEGIEGLTGSGS 5134
Qy 193 ----- 192
Db 5135 VASGRVAYTLGLEPAAVTVDTACSSSLVALHLAVQALRKGEVDMALAGGVVMWSTPSTFV 5194
Qy 193 -----TKVNDV 198
Db 5195 EFSRQRGLAPDGRSKSFSTADGTSWSEGVGLLVERLSDARRKGHRILAVVRGTAVNQD 5254
Qy 199 SQTIFYLGLTA----- 209
Db 5255 GAS---SGLTAPNGPSQORVIRRALADARLTTSDDVVVEAHGTGTRLGDPFIEAQAIVATY 5311
Qy 210 ----- 209
Db 5312 GQGRDGEQPLRLGSLKSNIGHTQAAAAGVSGVIMVQMMRHGVLPKTLHVKEKPTDQVDWSA 5371
Qy 210 ----- 209
Db 5372 GAVELLTEAMDWPDKDGLRRAAVSSFGVSGTNAHVLEBAPAAEETPASEATPAVEPS 5431
Qy 210 ----- 209
Db 5432 VGAGLVPMVLSAKTPAALDAQITGRLLAASFASQRTDAADPGAVARVLAGRABFEHRAVVL 5491
Qy 210 ----- 209
Db 5492 GTQDDFAQALTAPEGLIRGTPSDVGRVAFVFPQGTQWAGMGAEILLDVSKFAAAMAEBC 5551
Qy 210 ---LLRYAQ----- 216
Db 5552 ESALSRYTWSLEAVVROAPGAPTILERVVQVPTFAMVSLAKVWQHIGVTPQAVVGHS 5611
Qy 217 ----- 216
Db 5612 QGEIAAAVYAGALTDDAARVVVTLRSKSIAAHLAKGGMI SLALSEEATRQRIENLHGLS 5671
Qy 217 ----- 216
Db 5672 IAAVNGPTATVVGDPDTQIQELAQACEADGVRARIIPVDYASHSAHVETIESELAEVLAG 5731
Qy 217 -----RNCYH----- 221
Db 5732 LSPRTPEVPFFSTLEGAWITEPVLDTGYWRNLRHRVGFAPAVETLATDEGTFHEIVSA 5791
Qy 222 ----- 221
Db 5792 HPVLTMTLPTETVTLGTLRREGQGERLVTSLAEAWTNGLTIDWAPVLPTATGHHPELPT 5851
Qy 222 ----- 221
Db 5852 YAFQRRHYLHDSPAVQGSVQDSWRYRIDWKRLAVADASERAGLSGRVLVVPEDRSAEA 5911
Qy 222 ----- 221
Db 5912 APVLAALSGAGADPVQLDVSPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHPGH 5971
Qy 222 -----SFLVNA----- 228
Db 5972 PAPFTRGTGATLTLVQALEADAGVAAPLWCVTTHGAVSVGRADHVTSPAQMWGMGRVAAL 6031
```

QY 229 -----MSRNLFRVPKYNGTK 244  
Db 6032 EHPERWGLIDLPDADRALDRMTTVLAGTGEDQVAVRASGLLARLRVRLSPAHGTA 6091  
QY 245 ----- 244  
Db 6092 SPWQADGTVLVTCABEPAAAEARRLARDGAGHLLHTTPSGSEGAEGTSGAEDSGLA 6151  
QY 245 ----- 244  
Db 6152 GLVAELADLGATATVTCDLTDAEAAARLLAGVSDAHLPLSAVLHLPTVDSEPLAATDAD 6211  
QY 245 ----- 244  
Db 6212 ALARVVAKATAHLDRLLREAAAGRPVLVLFSSVAIIMGAGQAGYAAGTAFDA 6271  
QY 245 ----- 244  
Db 6272 LACQHRADGPTVTSVAWSPEWGSRVTEGATGERLRLGLRPLAPATALTALDTHLGHGT 6331  
QY 245 ----- 244  
Db 6332 AVTIADVMSFAPGFTTARPGTLLADLPEARALDEQQSTTAADTVLSRELALGTAE 6391  
QY 245 ----- 244  
Db 6392 QORRMQELVREHLAVVLNHPSPPEAVDTGRAFRDLGFDLSLTAVELNRLKNATGLALPATL 6451  
QY 245 ----- 244  
Db 6452 VFDYPTPTLAEFLAELIGEQAAGEQLPVDGGVDPEVAIVGMACRLPGGVASPEDLM 6511  
QY 245 ----- 244  
Db 6512 RLVAGGEDAISGPPQDRGWDVEGLYDPPDASRTYCRAGGLDEAGEFDADFFGISPRE 6571  
QY 245 -----LKNTRKJK----- 253  
Db 6572 ALAMDPORLLETWEAVEDAGIDPTSLQQQGVGVFAGTNGPHYBPLLRNTAEIDLEGYV 6631  
QY 254 ----- 253  
Db 6632 GTGNAASIMSGRVSYTLGLEGPATVDTACSSSLVALHLAVQALRKGEGLALAGGVTM 6691  
QY 254 ----- 253  
Db 6692 STPTTFEFSRQGLAEDGSKAFASADGFGPAEGVGMLLVERLSDARRNGHRVLAVR 6751  
QY 254 ----- 253  
Db 6752 GSAVNDGASGLTAPNGPSQQRVIRRALADARLTADVDVVEAHGTGRLGDPIEAQAL 6811  
QY 254 ----- 253  
Db 6812 IATYQGRDTEQPLRLGSLKSNIGHTQAAGVSGIIMVQAMRHGVLPKTLHVDPSDQI 6871  
QY 254 -----RKQ-----APVKE----- 261  
Db 6872 DMSAGTVELLTEAMDPKQEGGLRRAAVSSFGISGTNAHIVLEAPVDEDPADAPDEPSVG 6931  
QY 262 -----QFEKKA----- 267  
Db 6932 GVVPLVLSAKTPAALDAQIGRLAASFASQGRTTDAADPGAVARVLAGGQAQFEHRAVALGTG 6991  
QY 268 ----- 267  
Db 6992 QODLAAALAAPEGLVRGVASGRVAFVFPQGQTONAGMGAELLDVSKFAAAMACEAA 7051  
QY 268 ----- 267  
Db 7052 LAPYDWSLEAVVRQAPGAPTLELRLVDVQVTFVAVMVSLAKVWQHGHGVTPOAVVGHSGQE 7111  
QY 268 ----- 267

Db 7112 IAAAVAGALSDDAARVVVTLRSKSIHAHQGGMLSLALSEAAVVERLAGDGLSVAA 7171  
QY 268 -----KKT 270  
Db 7172 VNGPTATVSGDPTQIQELAQACEADGVARIIIPVDYASHSAHVETIESELADVLAGLSP 7231  
QY 271 QSTTTPYFS----- 279  
Db 7232 QTPQVFPFSTLEGAWITEPALDGGVYRNLRHVRGFPAPAVETLATDEGTFHFVEVSAHPV 7291  
QY 280 -----YTTS-----AALNVTNV-----TYSIT 297  
Db 7292 LTMALPETVTGLTRRDNGGQHRLTSLAEAWANGLTVDWASLLPTTTTHDPLDYAFQ 7351  
QY 298 T----- 298  
Db 7352 TERYWQPDLISAAGDITSAGLGAEBHPLIGAVALADSDGCLLTGSLSLRTHPWLADHAV 7411  
QY 299 ----- 298  
Db 7412 AGTVLLPGTAFVELAFRAGDQVCDLVEBLTLDAPLVLPRRGAVRVQLSVGASDESGRRT 7471  
QY 299 -----AARRVSTGIA----- 309  
Db 7472 FGLYAHPEAPGEAEWTRHATGVLAAADRTAPVADPEAMPPEGAEPVDVDGLYERFAAN 7531  
QY 310 ---YRP----- 312  
Db 7532 GYGYGLFGVGVWRGDEVFADVALPAEVAGAEGRFGLHPALLDDAAVQAAGAGRGVR 7591  
QY 313 -----DSS----- 315  
Db 7592 RGHAAAVRLERDLLYAVGATALRVLAPAGPDTVSVSAADSSQGPVFAADSLTVLPVDPA 7651  
QY 316 ----- 315  
Db 7652 QLAAFSDPTLDALHLEMTAWDGAQAALPGAVVLGGDADGLAAALRAGGTEVLSFPDLTD 7711  
QY 316 ----- 315  
Db 7712 LIVEAVDRGETPAPATVVLVACPAAGPDGPEHVREALHGLMALMQWLADERFTDGLRLVLT 7771  
QY 316 ----- 315  
Db 7772 RDAVAARSGLRSTGQAAMVGLGRSAQTESPGRFVLLDLAGEARTAGDATAGDGLTTGD 7831  
QY 316 ----- 315  
Db 7832 ATVGTTSGDAALGSALATALGSPEQLALRDGALLVPLRLARAAAPAAADGLAAADGLAAL 7891  
QY 316 ----- 315  
Db 7892 PLPAAPALMRLBPGTDGSLSLTAAPGDAETLAPEPLGPGQVRIATRATGLNFRDVLIAL 7951  
QY 316 ----- 315  
Db 7952 GMYPDPMALMGTEGAGVVTATPGVTHLAPGDRVMGLLSGAYAPVVVADARTVARMPEGWT 8011  
QY 316 -----FMKSTMATOLRDLA----- 329  
Db 8012 FAQGASVPVVFULTAVYA--LRDLADVKGPERLLVHSAAGVGVMAAVOLARHMGVEVHGTA 8069  
QY 330 ----- 329  
Db 8070 SHGKWDALRALGLDDAHIASSRTLOFESAFRAASGGAGMDVVLNLSLAREFVDASLRLGP 8129  
QY 330 ----- 329  
Db 8130 GGRFVEMGKTVDYDAERVAADHPGVGRAFDLGEAGPERIGEMLAEVIALFEDGVLRLHP 8189  
QY 330 ---TW----- 331  
||

Db 8190 VTTWDRRARDAFRHVSQARHTGKVLTMPISGLDPEGTVLLTGCTGALGGIVARHVVGWE 8249  
Qy 332 ----- 331  
Db 8250 GVRLLLVSRGTDAPGAGELVHEALGADVSAACDVADREALTAVLDSIPAHEPLTA 8309  
Qy 332 -VYTT----- 335  
|:|  
Db 8310 VVHTAGVLSGTLPSMTAEDVEHVRPKVDAAFLDLBTLSTPGVDLAAAFVWFSSAAVFG 8369  
Qy 336 ----- 335  
Db 8370 GAGQAYAAANATLDALAWRRRTAGLPALSLGMLWAETSGMTGSLDTRSLRARGAT 8429  
Qy 336 ----- 335  
Db 8430 PMDSELTSLDDAAMRDPALVPIALDVAALRAQQRDGMPLLSGLTRGSRVGGAPVN 8489  
Qy 336 ----- 335  
Db 8490 QRRAAAGGAGEADTDLGRLAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRD 8549  
Qy 336 -LRYRQN----- 341  
|:|  
Db 8550 TGFDSLTAVELRNLNAATGLRLPATLVFDHPTPGELAGHLDELATAAGGSWAEGTSG 8609  
Qy 342 ----- 341  
Db 8610 DTASATDRQTTAALAELELDLEGVLASLAPAAGRPPELAARLALAAALGDDGDATDLDE 8669  
Qy 342 ----- 341  
Db 8670 ASDDDLFSFDKELGSDSFMANNEDKLRYLKRVTAELOQNTLRLEIEGRTHPEVAIVG 8729  
Qy 342 -----PFC----- 344  
|:|  
Db 8730 MACRLPGVASPEDLMQLVAGDGAISEFPQDRGWDVEGLYDPPDASGRTYCRSGGFLH 8789  
Qy 345 ----- 344  
Db 8790 DAGEFDADFGISPREALAMDPOQLSLTTTAWAIESAGIDPTALKSGSLGVFVGWHGTG 8849  
Qy 345 ----- 344  
Db 8850 YTSGOTTAVQSPELEGLHVSAGALFSGRIAYVLGTGDPALTVDTCSSSLVALHLAVQ 8909  
Qy 345 ----- 344  
Db 8910 ALRKGECMALAGVTVPFNADLFVQFSRQRLAADGRSKAFATSDGFGPAEGAGVLLV 8969  
Qy 345 ----- 344  
Db 8970 ERLSDARRNGHRILAVVSGAVNQDGSNGLTAPHGPSQQRVIRRALADARLAPGDVDVV 9029  
Qy 345 ----- 344  
Db 9030 EAHGTGRLGDPTEAQAALITYQEKSEBQPLRLGALKSNIGHTQAAAGVAGVIRMVQAM 9089  
Qy 345 -----RNRNAVSEF--MKNTHV- 362  
|:|  
Db 9090 RHCLLPKTLHVDPSQIDWSAGTVELLTEAVDWPKEQDGLRRAAVSFGISGTHAVV 9149  
Qy 363 ----- 362  
Db 9150 LEEAPAVEDSPAVEPPAGGVWPVSAKTPAALDAQIGLAAVADGRDTPDPAVARAL 9209  
Qy 363 ----- 362  
Db 9210 VDSRTAMEHRAVAGDSREALRDLRMPGLVRGTSSDVRVAFVFPQGTQWAGMGAE 9269  
Qy 363 -----LIRNE----- 367  
|:|  
Db 9270 LDSSPEFAASMAECETALSRYVDWSLEAVVRQEPGAPTILDRVDVVQVPTFAVMVSLAKW 9329

Qy 368 -TPYTIYG-----TLD-----MSSLYYNE 385  
|:|  
Db 9330 QHHGITTPOAVGVHSGQEIAAAVYAGALTLLDAAVVVTLRSKSIAAHLACKGGMISLALDE 9389  
Qy 386 ----- 385  
Db 9390 AAVLKELSDFDGLSVAAVNGPTATVVGSDPTQIEELARTCEADGVPARIIPVDYASHSRQ 9449  
Qy 386 ----- 385  
Db 9450 VEIIEKELAEVLAGLAPOAPHVPFFSTLEGITWITEPVLDTYWRNLRRHVRGFAFAPAVETL 9509  
Qy 386 -TMFVE----- 390  
|:|  
Db 9510 AVDGFTHFIEVSAHPVLVMTLPETVTGLTLRREGQGBRLVTSABAWANGLTIDWAPI 9569  
Qy 391 ----- 390  
Db 9570 LPTATGHPELPTAFQTERFQLQSSAPTSAAEDWRYRVEWKPLTASQOADLSGRMIVAV 9629  
Qy 391 ----- 390  
Db 9630 GSEPEAELGALKAAAEVDVLEAGADDDREAAARLTALTITGDTGFTGVVSLDLLDPQV 9689  
Qy 391 ----- 390  
Db 9690 AMVQALGDAGIKAPLWSVTQGAHSVGRLDTPADPDRAMLWGLGRVVALEHPERWAGLVDL 9749  
Qy 391 ----- 390  
Db 9750 PAQPDAAALHLVTALSGATGSDQIAIRTTGLHARRLARAPLHGRRPTRDMQPHGTVLIT 9809  
Qy 391 -----NKTASDSNKT----- 400  
|:|  
Db 9810 GGTGALGSHAARWMAHGAHEHLLVSRSGEQAPGATQLTAELTASGARVTIAACDVAPH 9869  
Qy 401 ----- 400  
Db 9870 AMRTLLDAIPAEPTPLTAVVHTAGAPGDPDLVTGPDGIARILGAKTSGAEVLDLLRGT 9929  
Qy 401 ----- 400  
Db 9930 LDAFVLYSSNAGVMSGQGVVAAANAHLDAARRRARGETATSVANGLWAGDGMGRGA 9989  
Qy 401 -TPTSP-----SMGFQRTFI----- 414  
|:|  
Db 9990 DDAYWQRGIRPMSPDRALDELAKALSHDETFFVAVADVWERFAPFTVSRPSSLDDGVP 10049  
Qy 415 ----- 414  
Db 10050 EARQALAAPVGAPAGDAAVAPTQSSSALAATALPEPERPALLTLVRTHAAAVLGHSS 10109  
Qy 415 ----- 414  
Db 10110 PDRVAPGRAFTELGFDLSLTAVQLRNQLSTVVGNNLPATTVPDHTPAALAAHLHEAYLAP 10169  
Qy 415 -----DP----- 416  
|:|  
Db 10170 AEPAPTWEGRVRRALAEPLDLRLDAGVLTDLVRLTGIEPEPGSGSDGGAADPGABPE 10229  
Qy 417 ----- 416  
Db 10230 ASIDDDAEALIRMALGPRTMTSSNEQLVDALRASLKENELKESRRRRDRQEPMAI 10289  
Qy 417 -----LWD-----YLSLLF 426  
|:|  
Db 10290 VGMSCRFAGGIRSPEDLMDAVAAGKOLVSEVPEBERGMDIDSLYDPVGRKGTTVVRNAAF 10349  
Qy 427 LDETRNF----- 433  
|:|  
Db 10350 LDDAAGDAAPFGGISPREALAMDPOQRQLLEASWEVERAGIDPASVRGTDVGVVGGY 10409

QY 434 ----- 433  
Db 10410 QDYAPDIRVAPECTGGYVVTGNSSAVASGRIAYSLGLEPAVTVDTACSSSLVALHLAK 10469  
QY 434 ----- 433  
Db 10470 GLRNGDCSTALVCGVAVLATPGAFIEFSSQQAADGRKYGFAAADGLANGEGVAVLLL 10529  
QY 434 ----- 433  
Db 10530 ERLSDARRKGRHVLAVRGSAINQDGSNGLTAPHGSSQQLIRQALADARLTSSDDVV 10589  
QY 434 ----- 433  
Db 10590 EHGTCGRLGDPTEAQALLATYQGRAPQGLRLGLTKSNIIGHTQAAGVAGVVKMVQAL 10649  
QY 434 ----- 433  
Db 10650 RHGVLPKTLHVEPTDQVDSAGSVELLTEAVDWPBPGRLRAGVSAGVGGTNAHVVL 10709  
QY 434 ----- 433  
Db 10710 EEAPAVEESPVEPPAGGVVWPPVSAKTSALDAQIGQLAAVAEDTVDPAVARALV 10769  
QY 434 ----- 433  
Db 10770 DSRTAMEHRAVAGDSREALDALRMPEGLVRGTVDPCRFAVFPQGTQWAGMGAELL 10829  
QY 434 ----- 433  
Db 10830 DSSPEFAAAMAECEETALSPYVDSLEAVVRQAPSAPTLRDVVVQVPTFAMVMSLAKVWQ 10889  
QY 434 ----- 437  
Db 10890 HHGITEAVIGHSGEIAAYVAGALTLDAAARVTLRSKSAIAHLAKGGMISLALSEE 10949  
QY 438 ----- 437  
Db 10950 ATRQRIENLHGLSIAAVNGPTATVVSQDPTQIQELAQCEADGIRARIIPVDYASHAHV 11009  
QY 438 ----- 437  
Db 11010 ETIENELADVLAGLSPTQVPPFFSTLEGTWITEPALDGGYVYRNLRHVRGAPAVETLA 11069  
QY 438 ----- 441  
Db 11130 LPATGALSPAVPDLPTTYPAFQHRYSYWIWPAGPGEAPAHATASGREAVETGLWPGAEIDL 11189  
QY 442 ----- 444  
Db 11190 EGRRSVLAWMVROAASVLRCDSPPEVPVDRPLREIGFDSLTVDFNRVNRNLTLGLQP 11249  
QY 445 ----- 451  
Db 11250 PTVVFOHPTVALAERISDELAERNWAVAEPSDHEQAEKEKAAAPAGARSGADTGAGCM 11309  
QY 452 ----- 451  
Db 11310 FRALPQAVEDDRYGEFLDLAEASAFRQFASPEACSERLDPVLLAGGTDRAEGRVL 11369  
QY 452 ----- 457  
Db 11370 VGCTGAANGPHEFLRLSTSFQERDFLAVPLPGYGTGTGTALLPADLOTALDAQAR 11429  
QY 458 ----- 457  
Db 11430 AILRAAGDAPVLLGHSGGALLAHELAFRLERAHGAPPAGIVLVDPPPHQEPIEWNR 11489  
QY 458 ----- 457

Db 11490 QLGEGLFAGEBPMGDARLLAMGRYARFLAGPRPGRSSAPVLLVLRASEPLGDWQBERGDW 11549  
QY 458 -----NSLW 461  
Db 11550 RAHWDLPHTVADVDPGDHFTMRDHAPAVAEAVLSWLDATIEGEGAGKMTDRPLNVDSGLW 11609  
QY 462 ----- 461  
Db 11610 IRRFHPAPNSAVRLVCLPHAGGSASYFFRFBELHPSVEALSVOYPCRQDRRAEPCLESV 11669  
QY 462 -----WW 463  
Db 11670 BELAEHVVAATEPFW 11684  
RESULT 5  
US-09-108-006C-1  
; Sequence 1, Application US/09108006C  
; Patent No. 6524613  
; GENERAL INFORMATION:  
; APPLICANT: Steer, Clifford J.  
; Kren, Betsy T.  
; Bandyopadhyay, Paramita  
; Roy-Chowdhury, Jayanta  
; TITLE OF INVENTION: Hepatocellular Chimeraplasty  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kimeragen, Inc.  
; STREET: 300 Pheasant Run  
; CITY: Newtown  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 18940  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/108,006C  
; FILING DATE: 30-Jun-1992  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/054,288  
; FILING DATE: 30-APR-1997  
; APPLICATION NUMBER: 60/054,837  
; FILING DATE: 03-AUG-1997  
; APPLICATION NUMBER: 60/064,996  
; FILING DATE: 10-NOV-1997  
; APPLICATION NUMBER: 60/074,497  
; FILING DATE: 12-FEB-1998  
; APPLICATION NUMBER: PCT US 98/08834  
; FILING DATE: 30-APR-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Friebe, Thomas  
; REGISTRATION NUMBER: 29258  
; REFERENCE/DOCKET NUMBER: 7991-015-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-504-4444  
; TELEFAX: 215-504-4545  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4563 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-108-006C-1

Beet Local Similarity 4.7%; Pred. No. 0.0015; Matches 213; Conservative 74; Mismatches 152; Indels 4091; Gaps 63;		
Qy	12	PKMFVLISIFLLV-  : :: :
Db	6	PALLALLALLALLLAGARAEEMLENSLVCPKDRFVKHLRKYYTYNEAESSGVP 65
Qy	26	-----SPINCKV  : :: :
Db	66	GTADRSATRINCKVELEVPLQCSFILKTSQCILKEVYGFNPEGKALLKTKTNGSEFAAA 125
Qy	33	----- 32
Db	126	MSRYELKLAIPGEQVFLYPEKDEPTYIILNIRGIIISALLVPPPETEBAKQVFLDVTYGN 185
Qy	33	----- 32
Db	186	CSTHTVTRKGNVATEISTERDLGQDRFKPIRTGISPLALIKGMRPLSTLISSQSC 245
Qy	33	----- 32
Db	246	QYTLDAKRKHVAEAIKEOHLFLPFSYKQKYGMVAQVTTQLKLEDPKINSRPFEGGTCK 305
Qy	33	----- 32
Db	306	MGLAFESTKTSPPKQAEAVLTVQELKKLTISEQNIQRANLKNKLTVELRGLSDEAVTS 365
Qy	33	----- 32
Db	366	LLPOLIEVSPITLQALVQCGPOQSTHLOWLKRKHANPLLDVITYLVALIPEPSAQ 425
Qy	33	----- 32
Db	426	LREIFNMARDQRSRATLYALSHAVNNYHKNPTGTOQLLDIANYLMQIQDDCTGDEDT 485
Qy	33	----- 32
Db	486	YLILRVGNMGOTMEQLTPELKSSILKCVQSTKPSLMIQKAAIQALRMBPDKQDEVL 545
Qy	33	----- 32
Db	546	QTFLLDASPGDKRLAAYLMLRSPSQADINKIVQLPWEQNEQVKNFVASHIANILNSEE 605
Qy	33	----- 32
Db	606	LDIQDLKKLVKEVLKESQLPTVMDFRKFSRNYQLYKSVISPLDPASAKIEGNLIIDPNN 665
Qy	33	-----MSKALYRNPW- 42 :: :: :
Db	666	YLPKESMLKTTLTAFGPASADLIEIGLEKGFEPTEALFGKQGFPPDPVSNKALY---WV 722
Qy	43	-----RGLVLS-  : :: :
Db	723	NGQVPDGVSKVLVDHYGTDDKHEQDMVNGIMLSVEKLIKDLKSKEVPEARAYLRILGE 782
Qy	49	KIGKYKLQDLK- :: : :
Db	783	ELGFASLHDLQLLGLKLLMGARTLQGIQMTGEVIRKSGKNDFFLHVIFMENAFELPTGA 842
Qy	66	-----LETTISTKYNVS- :: : :
Db	843	GLQLQISSGVIAPGAKAGVKLEVMANQAEIYAVKPSVSVFVTNMGIIIPDFARSQVQN 902
Qy	78	-----KQPVKNL-  : :: :
Db	903	TNFFHESGLEAHVALPKGLKFIIPSPKRPVKLLSGGNTLHLVSTTKTEVIPPIENRQS 962
Qy	90	-----PF-  : :: :
Db	963	WSVCKVFPGLNYCTSGAYSNASDTSASYPLTGDTFRLLELRPTGEIQYSVSATYEL 1022
Qy	99	----- 98

QY	216	-----	215	QY	387	-----	386
Db	2163	AKINFNEKLSQLOQYMIQFDYIKDSYDLHDLKIAIAIIDEIIEIKLSDEHYHVRNL	2222	Db	3238	SHDELPRTEIQIPGYTVPVVNVVEVSPTIEMSAFYVFPKAVSMPSFISILGSDVRVPSYTL	3297
QY	216	-----	215	QY	387	-----	386
Db	2223	VKTIHDLHLFIENIDFNKSGSSTASWIONVDTKYQIRIQIQEQKLOQLKRHIQNIHQHLA	2282	Db	3298	ILPSLELPVLHVPRNLKLSLPDPKELCTISHIFIPAMGNITVDYDFSKSSVITLNTNAELF	3357
QY	216	-----	242	QY	387	-----	395
Db	2283	GKLGKHIEADVRVLDQGTTFISFERINDVLEHVHVFVNILIGDPEVAEKINAPRAKVH	2342	Db	3358	NQSDIVAHLLSSSSSVIDALQYKLEGTTRLTRKRGKGLATALSLSNKFVEGSHNSTVSLT	3417
QY	243	-----	265	QY	396	-----	413
Db	2343	ELIEREYVDOQIQVLMKDLVELAHQYKLETKIQKLSNVLQOVVKIDYFEKLVGFIDDAVK	2402	Db	3418	TKRMEVSVATTTTQAQIPILRMNFQKQELNGNTKSKPTVSSSMFEKYDFNSMLYSTAKGAV	3477
QY	266	-----	265	QY	414	-----	413
Db	2403	KLNELSKTFIEDVNFKFLDMLIKKLSFDYHQFVDETNDKIREVTQRLNGEIQALELPQK	2462	Db	3478	DHKLSLESLSYFSIESSTKGDVKSLSREYSGGTIASEANTYLSKSTRSSVKLOQTSK	3537
QY	266	-----	268	QY	414	ID-----	415
Db	2463	AEALKPLBETKATVAVYLESQDTKITLIINWLQEALSASLAHMAKAFRETFLEDRDR	2522	Db	3538	IDDINNLEVKENFAGEATLQRIYLSLWEHSTKNHLQLEGLFPTNGEHTSKATLELSPQMS	3597
QY	269	-----	308	QY	416	-----	415
Db	2523	MYQMDIQBELQRYLSLVGVYSTLVYISDWMTLAAKNLTDPAEQYISQDWAKRMKALV-	2581	Db	3598	ALVQVHASQPSHFDPDILQGEVALNANTKNQIRKWKNEVRIHSGSFQSQVELSNDQEA	3657
QY	309	AYRPDSSF-----	324	QY	416	-----	421
Db	2582	-----	2637	Db	3658	HLDIAGSLEGHURFLKNIILPVYDKSLWDFLKDVTTSIGRRQHLRVSTAFVYTKPNGY	3717
QY	325	-----	328	QY	422	-----	421
Db	2638	NIKIPRSTPELTINTHIPSFTIDFVEMVKYIIRTIDQMLNSLQHPVDPDIYLRDLK	2697	Db	3718	SFSIPVKVLADKFIIPGLKLNLSVLMPTFHVPTDLQVPCKLDFREIYIYKKLRTS	3777
QY	329	-----	328	QY	422	-----	430
Db	2698	VEDIPLARITLPDRPLPEIAPEFIITPLNLDQVDPDLHPFOLPHISHTIEVPTFGK	2757	Db	3778	SPALNPLTLPVEKFPPEVDVLTYSQPEDSLIPFEITVPESQLTVSQFTLPKSVSDGIAA	3837
QY	329	-----	328	QY	431	-----	430
Db	2758	LYSILKIQSPLFLDANADIGNTTSANEAGIAASITAKGESKLEVLNFDFOANAQSNP	2817	Db	3838	LDLNNAVANKIADPELTIIVPEQTIEIPSIKFSVPAGIAIPSFQALTARFVDSPVYNAT	3897
QY	329	-----	328	QY	431	-----	434
Db	2818	KINPLAKESVKFSSKYLRTHEGSEMLFFPGNAIEGKSNVTASLHTEKNTLELSNGVIVKI	2877	Db	3898	WSASLKNKADYVETVLDSCTSVQFLEYELNVGLGTHKIEDGTFLASKTKGTFARHDSAE	3957
QY	329	-----	335	QY	435	-----	439
Db	2878	NNQLTDSNTKYPHKLNIPKLDPSQADLRNEIKTLKAGHIAWTSSGKSGMKWACPRFS	2937	Db	3958	YEEDGKYEGLOEWEGKAHLNIPKSPAPFTDLHLRYQDKKGIKSTSAASPAVGTVMGMDDED	4017
QY	336	-----	341	QY	440	-----	439
Db	2938	DEGTHESQISFTIEGPLTSFGLSNKINSKHLRVNQNLVYESGSLNFSKLEIOSQVDSQHV	2997	Db	4018	DFSKNFFYSSPOSSDPKKLTIPTKTELVRRESDEETQIKVNBEEAASGLLTSLKONVPA	4077
QY	342	-----	350	QY	440	-----	442
Db	2998	GHSVLTAKGWALFGEKGAFTGRHDAHLNGKVIQTLKNSLFTSAQPFETASTNNEGNLK	3057	Db	4078	TGVLYDYNNKYHWEHTGLTLREVSSKLRNLDHAEWVYQGAIREIDIDERFQKASGT	4137
QY	351	-----	364	QY	443	-----	442
Db	3058	VRPPLRTGKIDPLNNVALFLSPSAQAQASQVARSFNQYKYNQNFSGAGNENIMEAHVGI	3117	Db	4138	TGTYQEWKDKAQNLYQELLTQEQASFOGLKDNVDFGLVRVTQEFHMKVKHLIDSIDFL	4197
QY	365	RNET-----	375	QY	443	-----	442
Db	3118	NGEANLDFNLPIPIEMRLPYITITPPLKDFSLWEKTEKLEFLTKTTKQSFDSVKAQY	3177	Db	4198	NPPRFQPGKPGIYTREELCTMFIREVGTSLVQVSKVNGSEILPSYFQDLVITLPPFL	4257
QY	376	-----	386	QY	443	-----	442
Db	3178	KQKXHRHSITNPLAVLCEPISQISKFSRHRHFEKRNRNALDFVTKSYNETKIPDKYKAEK	3237	Db	4258	RXKHLIDVISMYRELLKDLSEKAEQVFKAIQSLKTTTEVLRLNLDLQFIFOLLIEDNIKQL	4317
				QY	443	-----	442

Db 4318 KEMFTYLYNIODEINTFNDYIPYVFKLLKENLCLNHLKFNFIQNELQASQEIQQI 4377  
QY 443 -----LTPP----- 442  
Db 4378 HOYIMALREBYFDSIVGWTVKYVELEBKIVSLIKNLLVALKDPHSEYIIVSASNFTSOLS 4437  
QY 443 -----LTPP----- 446  
Db 4438 SQVEQFLHRNIQVLSILTPDGKGKIEAELSATAQEIISQAIATKLIISDYHQQFRI 4497  
QY 447 -----EHRVAVNLSTSN 458  
Db 4498 KLQDFSDQLSDYVEKFAESKRLLDLSIGN 4527

RESULT 6  
US-09-538-092-842  
; Sequence 842, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratPatSeqFormatter Version 0.9  
; SEQ ID NO 842  
; LENGTH: 4563  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P04114  
US-09-538-092-842

Query Match 18.3%; Score 449; DB 4; Length 4563;  
Best Local Similarity 4.6%; Pred. No. 0.0015;  
Matches 209; Conservative 74; Mismatches 151; Indels 4101; Gaps 61;

QY 12 PKMFVLISFLAV----- 25  
Db 6 PALLALLALLALLLAGARAEEMLENVSLVCPKDATRFKHLRYTYNREABSSSGVP 65  
QY 26 -----SFINCKV----- 32  
Db 66 GTADSRATRNCKVEVLPQLCSFILKTSQCTLKEVYGFNPEGKALLKTKNSBEFAAA 125  
QY 33 ----- 32  
Db 126 MRYELKLAPEGKQVFLYPEKDEPTYIILNKRGIISALLVPPETBEAKQVFLDVTYGN 185  
QY 33 ----- 32  
Db 186 CSTHFTVTRKGNVATEISTERDLGQDRFKPIRTGISPLALIKGMRPLSTLISSQSC 245  
QY 33 ----- 32  
Db 246 QVTLDAKRKHVAEAIKCEQHLFLPFSYNNKYGMAQVOTLKLDETPKINSRFFGEGTKK 305  
QY 33 ----- 32  
Db 306 MGLAFESTKTSPPKQAEAVLKTQELKKLTISEQNIQRIANLFNKLVTELRLGLSDEAVTS 365  
QY 33 ----- 32

Db 366 LLPQLIEVSSPITLQALVOCGQPCSTHILQWLKRVHANPLLIDVVYLVVALIPEPSAQO 425  
QY 33 ----- 32  
Db 426 LREIFNMARDORSRATLYALSHAVNNYHKTNPGTQELLDIANLYLMEIQDDCTGDEDYT 485  
QY 33 ----- 32  
Db 486 YLILRVIGNMGQTMEOITPELKSSILKCVQSTKPSLMIKAAIQALRKMEPKDKQDEVLL 545  
QY 33 ----- 32  
Db 546 QTFLDASPGDKRLAAYLMLMRSPQADINKIVILPWEQNEQVKNFVASHIANILNSE 605  
QY 33 ----- 32  
Db 606 LDIQDLKLVKEALKESQLPTVMDPRKFRSNRYLYKSVSLPSLDPASAKIEGNLIFDPNN 665  
QY 33 -----MSKALYNRPW- 42  
Db 666 YLPKESMLKTTLTAFGFASADLIEIGLECKGPEPTLEALFGKQGFPPDSVNKALY--WV 722  
QY 43 -----RGLVLS----- 48  
Db 723 NGQVPDGVSKVLVDHFGYTKDDKHEQDMVNGIMLSVEKLIKDLKSKVEPEARAYLRLGE 782  
QY 49 ----- 48  
Db 783 ELGFASLHDLQLLGLKLLMGARTLOGIPOMIGEVIRKGSKNDFFLHYIPMENAFELPTGA 842  
QY 49 ----- 48  
Db 843 GLQLQISSSGVIAPGKAGVKLEAVANMQAELVAKSVSVFVTNMGIIPDFARSGVQMN 902  
QY 49 -----KIGKYK----- 54  
Db 903 TNFFHESGLEAHVALKAGKLFIIIPSPKRPVKLLSGGNTLHLVSTTKTEVIPPLIENROS 962  
QY 55 ----- 54  
Db 963 WSVCKQVFPGLNYCTSGAYSNASTDSASYPLTGTDRLELRPTGEIEQVSVSATVEL 1022  
QY 55 -----LDQLK-----BILRQLETTISTKYNVSKQPVKNLTMTNTE--FPQY----- 93  
1023 QREDRALVTLTKFVTOAGAKQTEATWTFKYN-----RQSMTLSEVOIPDPDVLGTIL 1077  
QY 94 -----YILAGPIQNYSIYIL----- 108  
1078 RVNDESTEKTSYRLTLDIONKKITEVALMGHLSCTDKERRKIKGVISIPRLQABARSEI 1137  
QY 109 ----- 108  
1138 LAHWSPAKLLQMDSSATAYGSTVSKRVAMHYDEKIEFENWTGNVDTKMTSNFPVDL 1197  
QY 109 -----WF----- 110  
1198 SDYPKSLHWANRLIDHRVPETDMTFRHVGSKLIVAMSSWLQKASGLPYTQTLQDHLNS 1257  
QY 111 -----DFVSTQ-----LR 118  
1258 LKEFNLQNMGLPDPHIPENLFLKSDGRVKYTLNKNLSKIEIPLPGGKSRDLKMLETVR 1317  
QY 119 KPQKY-----VYSQ-YNHNTA----- 132  
1318 TPALHFKSVGFHLPSREPVFTTIPKLYQLQVPLGLVLDLSTNVYSLNLYNWSASYSGN 1377  
QY 133 -----KTITP----- 137  
1378 TSDHFSLRARYHMKADSVDDLSTNVQSGGETTYDHKNTFTLSCDGLSRHKFLDSNIKF 1437  
QY 138 ----- 137  
1438 SHVEKLGNNPVSKGLLIFDASSWGPQMSASVHLDSKKQHLFVKEVKIDQFQVSSFYA 1497



QY 138 -----RPPCGRVP-----SMTCLSEM-----154  
DB 1498 KGTYGLSCQDPNTGRNGESNLFNSYLGQTNQITGRYEDGTLSTSTSLQSGIKN 1557  
QY 155 -----LNVSREN-----161  
DB 1558 TASLKYENVELTKSDNTGKYKNFATSNKMDMTFSKONALLRSEYQADYESURFFSLSG 1617  
QY 162 -----DTGB 165  
DB 1618 SLNSHGLELNADILGTDKINSAGAKATLRIGQDGI STSATTNLC SLLVLENELENAELGL 1677  
QY 166 QCGG-----NF-----171  
DB 1678 SCASMKLTNGRPREHNAKFLDGAALTELSGAYQAMILGVDSKNIFNFVKVQOGLK 1737  
QY 172 -----TTF 174  
DB 1738 LSNMMSGVAEMKFDHTNSLNIAGLSLDFSSKLDNIYSSDKFYKQTVNLQLOPYSLVTTL 1797  
QY 175 N-----PM-----177  
DB 1798 NSDLKYNALDLTNGKLRLEPLKLVAGNLKGAYQNEIKHIYAISSAALSASYKADTVA 1857  
QY 178 -----FNVPR-----183  
DB 1858 KVGQVEFSHRLNTDIAGLASAIDMTNYSNLSLHFNVSFVSMPTMTIDAHTNGNGL 1917  
QY 184 -----183  
DB 1918 ALWGEHTGQLYSKFLKABPLAFTFHDYKGSTSHLVSRSKISAALHKKVSALLPBAEQ 1977  
QY 184 -----WTKLVG-----PTKV-----195  
DB 1978 TGTWKLKQFNNEYSQDLDAYNTKKIGIVELTGRTLADLTLLDSPIKVPLLSEPINII 2037  
QY 196 -----NVDQTI-----YF-----204  
DB 2038 DALEMRDAVEKQOEFTIVAFVKYDKQVHSINLPFFTELQVEYFERNROTIIVVENVQR 2097  
QY 205 -----LG-----LTALLRYA-----215  
DB 2098 NLKHINIDQVRKYRAALAGKLPOQANDYLSNFNWERQVSHAKEKLTALTCKYRITENDIQ 2157  
QY 216 -----215  
DB 2158 IALDDAKINFNEKLSOLOQTMIOFDQYIKDSYDLHLDKTAIAIIDEIIEKLSLDEHYH 2217  
QY 216 -----215  
DB 2218 IRVNLVKTTHDLHLPFENIDFNKSGSSTASWIONVDTKYIRIQIQEKLQQLKRHIQND 2277  
QY 216 -----QRNCTHSPYLVNMSRNL-----FRVPKYING- 242  
DB 2278 IQHLAGLKQHIBAUDRVLLDQGTGISFERINDVLEHVGHFVINLIGDFEVAEKINAF 2337  
QY 243 -----TKLNTMRKLKR--KOAPVKEQPEK-----265  
DB 2338 RAKVHELIEREYVDQIQVLMKDLVELTHQYKLUKTIQKLSNVLOQVKIKDYFEKLVGFI 2397  
QY 266 -----265  
DB 2398 DDAVKKLNLSFKTFIEDVKNPLDMLIKKLSFDYHQFVDETNDKIREVTQRLNGBIQAL 2457  
QY 266 -----KAK-----268  
DB 2458 ELPQKAEALKFLEETKATVAVYLESIQDTKITLIINWLOEALSSASLAHMAKAFRETLE 2517  
QY 269 -----KTOSTTTPYES-YTTSALNVTTNV-TYSITTAAREV 303  
DB 2518 DTRDRMYQMDIOELORYLSLVGQVYSTLVITYISDWTIAARNLDTFASQYISQDWAKRM 2577

QY 304 STSTIAYRDPDSF---MKSIMATO-----324  
DB 2578 KALV-----EQGFTVPEIKTILGTMPAFEVSLQALQKATQTFPDFIVPLTDLRIPSQIN 2632  
QY 325 -----324  
DB 2633 FIDLKNIKIPSRFSPFTILNTFHIPSTIDFVEMKVKIIRTIQONQSELOWPVPDIY 2692  
QY 325 LRDL-----328  
DB 2693 LRDLKVEDIPLARITLDPFRLPEIAIPEFIIPTLNLDFQVDPDLHIPFQLPHISHTIEV 2752  
QY 329 -----328  
DB 2753 PTFGLKYLKIQSPLFTLDANADINGTTSANEAGIAASITAKGBSKLEVLNDFQANA 2812  
QY 329 -----328  
DB 2813 QLSNPKINPLALKESVKFSKYLRTEHSEMFLFPGNAIEGKSNVTASLHTEKNTLELSNG 2872  
QY 329 -----ATWVYT 334  
DB 2873 VIVKINNQLTLDSTNKYPHKLNI PKLDFSSQADLRNEIKTLKAGHIAWTSSGKGSWKA 2932  
QY 335 T-----LRYQN-----341  
DB 2933 CPRFSDGTHESQISFTIEGPLTSFGLSNKINSKHLRVNQNLVYESGSLNFSKLEIQSQV 2992  
QY 342 -----PF-CBPSRNR 350  
DB 2993 DSQHVHVSULTAKGMALFOEGKAETGRHDAHLNGKIVIGTLKNSLFFSQPFEITASTNN 3052  
QY 351 -----TAVSEFMKN-----359  
DB 3053 EGNLKVREPLRLTGKIDFLNNYALFLPSAQASQVSAFNOYKYNQNSAGNENIME 3112  
QY 360 THVLIRNET-----PYTIYGT-----375  
DB 3113 AHVINGEANLDFNLPIPTIPEMRLPYTIITPPPLKDFSLWEKTLKBEFLKTTKQSFDSL 3172  
QY 376 -----LDMSSLYNET-----386  
DB 3173 VKAQYKKNKGRHSITNPLAVLCEFIQSISKSPDRHPEKRNALDPVTIKSYNETKIKPKDK 3232  
QY 387 -----386  
DB 3233 YKAEKSHDELPTFQIPGYTVPVNVVSPFTIEMSAFCYVFPKAVSMPSPSLGSDVRV 3292  
QY 387 -----386  
DB 3293 PSYTLILPSLELPVLHVPRNLKSLPHPKELCTISHIPAMGNITYDFSKSSVITLNT 3352  
QY 387 -----MPVE---NK 392  
DB 3353 NAEFLNQSDIVAHLLSSSSSVIDALQYKLEGTTLTRKGLKLATALSLSNKFVEGSHNS 3412  
QY 393 TAS-----DSNKTPT-SPSMGFORTF-----413  
DB 3413 TVSLTTKNMEVSAKTTKAEIPILRMNPQBELNGNTKSKPTVSSSMEPKYDFDNSSMLYST 3472  
QY 414 -----413  
DB 3473 AKGAVDHKLSELSLTSYFSIESSTKGVKSVLSREYSGTIASEANTYLSKSTRSSVKL 3532  
QY 414 -----ID-----415  
DB 3533 QGTSKIDDIWNLEVKENFAGRATLORIYSLWEHSTKNHLOLEGLFTNGEHTSKATLELS 3592  
QY 416 -----415  
DB 3593 PQMSALVOVHASQPSSSHDPDLQGEVALNANTKNQIRWKNVEPRIHSGSPSQSVELSN 3652  
QY 416 -----PLWDYL-----421

Db 3653 DQKALHDIAGSLGHLRFLKNILPVYDKSLMDFLKDVTTSIGRRQHLRVSTAFVYTK 3712  
QY 422 -----  
Db 3713 NPNYGSFSPVKVLADKFIITPGLKLDNLNVLMPTFHVFTDLQVPSCKLDFREIQIYK 3772  
QY 422 -----  
Db 3773 KLRTSSPALNLPTLPEVKFPEVDVLTYSQSPEDSLIPFFFIITPESQLTVSQTLPSKVS 3832  
QY 431 -----  
Db 3833 DGIAALDLNAVAKIADFPETIIVPEQTIEIPSIKFSVPAGIVIPSFQALTARFEVDSP 3892  
QY 431 -----  
Db 3893 VYNATWSASLKNKADYVETVLDSTCSVTFLEYELNVLGHKIEDGTLASKTKGTLAHR 3952  
QY 432 NFS-----  
Db 3953 DFLSAEYEDCKFGLQEWGKAHLNLIKSPAFDHLRYQDKKIGISTASAAPAVGTVMGD 4012  
QY 440 -----  
Db 4013 MDEDDDFSKWPFYSPQSPDKLTIFKTELVRSEDEETQIKVNWEEEAASGLTSLKD 4072  
QY 440 -----  
Db 4073 NVPKATGVLYDYNNKYHWEHTGTLREVSSKLRNLQNAEWYOGAIRIDIDVRFOK 4132  
QY 443 -----  
Db 4133 AASGTTGTQEWKDKAQLYQELLTOEGQASFOGLKDNVDPGLRVTKQPHMKVKHLIDS 4192  
QY 443 -----  
Db 4193 LIDFLNPPRPQPKPGIYTBELCTMTFREVGTLSQVYKVHNGSEILFSYFDLVIT 4252  
QY 443 -----  
Db 4253 LPPELRKHLIDVISMYRELLKDSKAEQVFKAIQSLKTEVLRLNLDLLOFIFOLIED 4312  
QY 443 -----  
Db 4313 NIKQLKEMKFTYLINIYQDEINTIFNDYIPYVFKLKENLCLNLHKNFPIQNELQESQ 4372  
QY 443 -----  
Db 4373 ELQIHQYIMALREYFDPISVGTWKYVELEBKIVSLIKNLILVALKDFHSEYIVSASNF 4432  
QY 443 -----  
Db 4433 TSOLSSQVQFLHRNTQEVLSILTDPDGKGKEKIAELSATAEIKSQAIAATKKIISDYH 4492  
QY 447 -----  
Db 4493 QQPRYKLDQSDQSDYVEKFAESKRLDLSIQN 4527

RESULT 7

US-09-180-422B-27  
; Sequence 27, Application US/09180422B  
; Patent No. 6444644  
; GENERAL INFORMATION:  
; APPLICANT: BRUCKDOERFER, KARL R  
; ETTLELAIE, CAMILLE  
; TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED  
; FROM APOLIPROTEIN B-100  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE, P.C.  
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR

CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/180,422B  
FILING DATE: 07-Dec-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B.J.  
REGISTRATION NUMBER: 36663  
REFERENCE/DOCKET NUMBER: 117-268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 7038164000  
TELEFAX: 7038164100  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 27:  
US-09-180-422B-27

Query Match 18.0%; Score 440; DB 4; Length 4536;  
Best Local Similarity 4.8%; Pred. No. 0.003;  
Matches 213; Conservative 76; Mismatches 156; Indels 3970; Gaps 64;  
QY 1 MGKKE----- 5  
Db 99 MSRYELKLAIPGKQVFLYPEKDEPTYILNKRGIISALLVPPEEAKQVLFDTVYGN 158  
QY 6 ----- 5  
Db 159 CSTHTVTKRKNVATEISTERDLQCDRFPKPIRTGISPLALIKGWRPLSTLISSQSC 218  
QY 6 -----MMVRDVPK----- 13  
Db 219 QYTLDAKRKHVAEAIKESQHLFLPFSYKNKYGMVAQVTQTLKLEDTPKINSRFFGEGTK 278  
QY 14 -----MP----- 15  
Db 279 MGLAFESTKSTSPPKQAEAVLTKLQELKKLTISEQNIQRANLFNKLVTGLRLGLSDEAVTS 338  
QY 16 -----VLISIFLLVSEI----- 28  
Db 339 LLPQLIEVSSPITLQALVQCGPQCSSTHLOWLKRVRHANPLLDIVVTYLVALLIPEPSAQ 398  
QY 29 ----- 28  
Db 399 LEEIFNWARDQSRATLYALSHAVNNYHKNPTGTQELLDIANYLMEQIQDDCTGDEYDT 458  
QY 29 ----- 28  
Db 459 YLILRVIGNMGQMTPELTKSSILKCVQSTKPSLMIQAAIQALRKMPEKDKQDEVLL 518  
QY 29 ----- 28  
Db 519 QTFLLDDASPGDKRLAAYLMLMRSPSQADINKIVQLPWEQNEQVKNFVASHIANILNSEE 578  
QY 29 -----NCKV----- 32  
Db 579 LDIQDLKKLVKEVKESQLPTVMDPRKFSRVNQLYKSVSLPSLDPSAKIEGNLIFDPNN 638  
QY 33 -----MSKALYNRPW- 42  
Db 639 YLPKESMLKTTLTAFGFASADLIEIGLEGKGFPTLEALFGKQGFPPDSVKNALY---WV 695

QY 43 -----RGLVLS----- 48  
Db 696 NGQVPDGVSKVLVDHFGYTKDDRKHQDMVNGIMLSVEKLIKDLKSVEPEARAYLRILGE 755  
QY 49 ----- 48  
Db 756 ELGFASLHDLQLLGLKLLMGARTLQGIQOMIGEVIRKSGKNDPFLHYIIPMENAFELPTGA 815  
QY 49 ----- 48  
Db 816 GLQLOISSSGVIAPGAKAGVKLEVANQMAELVAKPSVSEFVTNMGIILIPDPARSQVQW 875  
QY 49 -----KIGYK----- 54  
Db 876 TNFFHESGLEAHVALKAGLKFIIPSPKRPVKLLSGNTLHLVSTTKTEVIPPLIENRQS 935  
QY 55 ----- 54  
Db 936 WSVCKQVFFGLNYCTSGAYSNASSTDSASYPTGTTRLELRPTGEIEQVSVSATYEL 995  
QY 55 -----LDQLK-----EILRQLETTISTKYNVSKQPVKNLTWNTB--PQY----- 93  
Db 996 QREDRALVTLKFTVQABGAKQTEATWPKYN-----RQSMILSSEVOIPDFDVLGTIL 1050  
QY 94 -----YILAGPIQNSYTYL----- 108  
Db 1051 RVNDESTEGKTSYRLTLDIQNKKITEVALMGHLSGCDTKEERKIKGVISIPRLQABARSEI 1110  
QY 109 ----- 108  
Db 1111 LAHWSPAKLLQMDSSATAYGSTVSKRVAMHYDEEKIBEFWNTGTNVDTKGMTSNPPVDL 1170  
QY 109 -----WF----- 110  
Db 1171 SDYPKSLHMYANPLLDHRRVPQDMDTPRHVSGKLIIVAMSSWLQKASGLPYTQTLQDHLNS 1230  
QY 111 -----DFYSTQ-----LR 118  
Db 1231 LKEFNLQNMGLPDFHIPENFLKSDGRVKVTLNKNLSKIEIPLPFGGKSSRDLKMLETVR 1290  
QY 119 KPACY-----VYSQ--YNHTA----- 132  
Db 1291 TPALHFKSVGFHLPSREFQVPTFTIPLKYQLQVPLLGVLDLSTNVYSNLYNNSASYSGGN 1350  
QY 133 -----KTIIF----- 137  
Db 1351 TSTDHFLSLRHYMKADSVVDLLSYNVQSGGETTYDHKFTPLSCDGLSRHKFLDSNIKF 1410  
QY 138 ----- 137  
Db 1411 SHVEKLGNNPVSKGLLIFDASSSWGPMQMSASVHLDSKKQHLFVKEVKIDGQFRVSSFYA 1470  
QY 138 -----RPPPCGRVP-----SMTCLSEM----- 154  
Db 1471 KGTYGLSCORDEPTNTRLNGESNLRFNSSYLQGTNQITGRYEDGTLSTLSTSLQSGIKN 1530  
QY 155 -----LNVSKRN----- 161  
Db 1531 TASLKYENYELTKSDTNKYNKPNFATSNKMDMTFSKQALLRSEYQADYESLRPFSLLSG 1590  
QY 162 -----DTGE 165  
Db 1591 SLNSHGLELNADILGTDKINGSAHKATLRIGQDGISTSATTNLKSLVLLENELNAELGL 1650  
QY 166 QCGG-----NF----- 171  
Db 1651 SGASMKLTJNGRPREHNAKFLSDGKAALTELSIGSAYQAMILGVDSKNIFNFKVSOEGLK 1710  
QY 172 -----TTF 174  
Db 1711 LSNDMGMSVAEMKFDHTNSLNIAGLSLDFSSKLDNIYSSDKFYKQTVNILOQOPYSLVTTL 1770

QY 175 N-----PM----- 177  
Db 1771 NSDLKYNALDLTNNGKRLRLEPLKHLVAGNLKAGYQNEIKHIYAISSAALSASYKADTVA 1830  
QY 178 -----FENVPR----- 183  
Db 1831 KVQGVFESHRLNTDIAGLASAIDMSTNYNSDSLHSNVSFRSVMAPFTMTIDAHTNGNGKL 1890  
QY 184 ----- 183  
Db 1891 ALWGEHTGQYSKFLKAEPLAFTSHDYKSGTSHHLVSRKSISSAALEHKVSALLTPAEQ 1950  
QY 184 -----WNTKLYVG-----PTKV----- 195  
Db 1951 TGTWKLKTFQNNNEYSQDLDAYNTKDKIGVBELTGRTLADLTLLDSPIKVPLLSEPINI 2010  
QY 196 -----NVDSTI-----YF----- 204  
Db 2011 DALEMRDAVEKPOEFTIVAFVKYDKNQDVHSINLPPFETLOEYFERNQTIIVLENVQR 2070  
QY 205 -----LG-----LTALLRYA----- 215  
Db 2071 NLKHINIDQFVKYRAALGKLQQANDYLSFNWERQVSHAKEKLTALTCKYRITENDIQ 2130  
QY 216 ----- 215  
Db 2131 IALDDAKINFNEKLSOLOTYMIQDQYIKDSYDLHDLKIAIANIIDEIIEKLSLDEHYH 2190  
QY 216 ----- 215  
Db 2191 IRVNLVKTIDHLHFIENIDFNKSGSSTASWTONVTKYQIRIQIOEKLQOLKRHTONID 2250  
QY 216 -----QRNCTSHFYLVNAMSRL-----FRPKYING----- 242  
Db 2251 IQHLAKLQHIIEADVRVLLDQLTGISFERINDVLEHVHGFVNILGDFEVAEKINAF 2310  
QY 243 -----TKLKNTRMKLKR--KQAPVKEQPEK----- 265  
Db 2311 RAKVHELIEREVDQIQVLMKCLVELAHQYKLETKIQKLSNVLOQVKIKDYPEKLVGFI 2370  
QY 266 ----- 265  
Db 2371 DDAVKKLNELSPKTFIEDVNKFLDMLIKKLKSFYHQFVDETNDKIREVTQRLNGEIQAL 2430  
QY 266 -----KAK----- 268  
Db 2431 ELPOKAEALKFLEETKATVAVYLESLODKITLIINMLQEALSSASLAHMKAKPRETLE 2490  
QY 269 -----KTQSTTPPYFS-YTTSAALNVTTNV--TYSITTAAARRV 303  
Db 2491 DTRDRMYQMDIQOELQRYLSLVGVYVSTLVTVYISDWTLAAKNLTDFAEQYSIQDWAKRM 2550  
QY 304 STSTIAYRPDSSF---MKSINATO----- 324  
Db 2551 KALV-----EQGTVPETIKTLGTWPAFVLSQALQKATQTPDFIVPLTDLRIPSVQIN 2605  
QY 325 ----- 324  
Db 2606 FKDLKNIKIPSRPSTPEFTILNTFHPSTIDFVEMKVKIIRTIDQMLNSELQWPVPDIY 2665  
QY 325 LRDL----- 328  
Db 2666 LRDLKVEDIPLARITLPDFRPLPEIAPFIIPTLNLNDFQVPLHIPPBQLPHISHTIEV 2725  
QY 329 ----- 328  
Db 2726 PTFGKLYSLTKIOSPLFTLDANADICNGTTSANEAGIAASITAKGESKLEVLNFDFOANA 2785  
QY 329 ----- 328  
Db 2786 QLSNPKINPLALKESVKSSKYLRTHEGSEMPLFPFGNAIEGKSNVTASLHTEKNTLELSNG 2845  
QY 329 -----ATWVYT 334



INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4536 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-079-030-1

Query Match 18.0%; Score 440; DB 4; Length 4536;  
Best Local Similarity 4.8%; Pred. No. 0.003;  
Matches 213; Conservative 76; Mismatches 156; Indels 3970; Gaps 64;

QY 1 MGRKE----- 5  
DB 99 MSRYELKLAPEKQVFLYPEKDEPTYILNKRGIISALLVPPETBEAKQVFLDTPVGN 158  
QY 6 ----- 5  
DB 159 CSHPTVTRKGNVATEISTERDLGQDRPKPIRTGISPLALIKGTRPLSTLISSQSC 218  
QY 6 -----MMVRDVPK----- 13  
DB 219 QYTLDAKRKHVAEAIKCEQHLPLFPSSNNKYGMAVQTOTLKLEDTPKINSRFFGEGTKK 278  
QY 14 -----MF----- 15  
DB 279 MGLAFESTKTSPPKQAEAVLKTQELKLTISEQNIQIRANLFNKLVTBLRGLSDEAVTS 338  
QY 16 -----VLISIFLLVSP----- 28  
DB 339 LLPQLIEVSPITLQALVQCGQPCSTHILQWLKRVHNPFLLDVTVYLVALIPEPSAQ 398  
QY 29 ----- 28  
DB 399 LREIFNWARDQRSATLYALSHAVNVYKTNPTGTQELLDIANYLMEIQDCTGDEDYT 458  
QY 29 ----- 28  
DB 459 YLILRVIGNMQTMEQLTPELKSSILKCVQSTKPSLMIKAAIQAALRMEPKDQOEVL 518  
QY 29 -----NCKV----- 32  
DB 519 QTFLDASPGDKRLAAYLMLMRSPSQADINKIVQILPWEQNEQVKNFVASHIANILNSEE 578  
QY 29 -----NCKV----- 32  
DB 579 LDIQDLKLVKEALKESQLPTVMDFRKSRNYQYKSVLSPLSDPASAKIEGNLIFDPNN 638  
QY 33 -----MSKALYNRPW- 42  
DB 639 YLPKESMLKTTLTAFGFASADLTIEIGLEGKGFPTLEALFGQGFPPDSVNRKALY---WV 695  
QY 43 -----RGLVLS----- 48  
DB 696 NGQVPGVSKVLVDHFGYTKDDKHEQDMVNGIMLSVEKLIKDLKSKEVPEARAYLRILGE 755  
QY 49 ----- 48  
DB 756 ELGFASLHDLQLLGLLLMGARTLQGI PQMIGEVIRKGSKNDFFLHYIFWENAFELPTGA 815  
QY 49 ----- 48  
DB 816 GLQIQISSGVIAPGAKAGVKLEAVANQAEVLVAKPSVSVEFTVMGIIIPDFARSQVMN 875  
QY 49 -----KIGYK----- 54  
DB 876 TNFFHESGLEAHVAKGLKFIIPSPKRPVKLLSGNTHLVSTTKTEVIPLPIENROS 935  
QY 55 ----- 54  
DB 936 WSVCKQVPGFLNYCTSGAVSNASSTDSASYPTGTDLRLLELRPTGIEIQSVSATYEL 995  
QY 55 -----LIDQLK-----EILRQLETTISTKYNVSKQPVKNLTWNTNTE--PFQY----- 93

996 QREDRALVDTLAKFVTOAEAGAKQTEATMTFKYN-----ROSMTLSSSEVQIPDFDVLGTIL 1050  
94 -----YILAGPTQNTSYITV----- 108  
1051 RVNDESTEKTSYRLTLDIQNKKIIEVALMGHLSCDTKEERKIKGVISIPRLQAEARSEI 1110  
109 ----- 108  
1111 LAHSPAKLLQWSSATAYSTVSKRVAWHYDEBKIEPEMNTGTNVDTKMTSNFPVDL 1170  
109 -----MF----- 110  
1171 SDYPSLHMYANRLLDHRVPETDMTRFHVGSKLIIVAMSSWLQKASGLPYTOTLQDHLNS 1230  
111 -----DFYSTQ-----LR 118  
1231 LKEFNLQNGMLPDFHLPENLFLKSDGRVKYTLNKSLEIPLPGGKSRDLKMLETVR 1290  
119 KPARY-----VYSQ-YNHTA----- 132  
1291 TPALHFKSVGFHLPSPREFQVPTTIPKLYQLQVLLGLVLDLSTNVYSNLYNMSASYSGN 1350  
133 -----KTITF----- 137  
1351 TSTDHFSLRARYHMKADSVVDLLSYNVQSGETTYDHRKMTFTLSCDGSURHKFELDSNIKF 1410  
138 ----- 137  
1411 SHVEKLGNNPVSKGLLIFDASSSWGPMQASVHLDSKKQHLFVKEVKIDGQFRVSSFYA 1470  
138 -----RPPPCGRVP-----SMTCLSEM----- 154  
1471 KGTYGLSCQDPNTGRLNGESNLNRSYLOQNTQITGRYEDGTLSLTSTSLQSGIKN 1530  
155 -----LNVSCKN----- 161  
1531 TASLKYENVELTKSDTNGKYKNFATSNQMDMTFSQNALLRSEYQADYESLRFFSLSG 1590  
162 -----DTGE 165  
1591 SLNSHGLELNADILGTDKINSGAHKATLRIGODGISTATSATTNLKCSLLVLENELNAELGL 1650  
166 QCG-----NF----- 171  
1651 SGASKMLTTNGRFRHNAKFLDGAALTELSGAYQAMILGVDSKNIENFKVSOEGLK 1710  
172 -----TTF 174  
1711 LSNMWSYAEMKFDHTNSLNTAGLSLDPFSKLDNIYSSDKPYKQTVNLQLOPYSLVTTL 1770  
175 N-----PM----- 177  
1771 NSDLKYNALDLTNGKLRLEPLKLVAGNLKAYQNNKIKHIVAISSAALSASYKADTVA 1830  
178 -----FNVPR----- 183  
1831 KVOGVESHRLNTDIAGLASAIDMTNYSNDSILHFSNVPRSVMAPFTMTIDAHTNGKGL 1890  
184 ----- 183  
1891 ALWGEHTQOLYSKFLKAEPLAFTFSDHYKSTSHHLVSRKSISSAALHKVSALLTPAQ 1950  
184 -----WNTKLYVG-----PTKV----- 195  
1951 TGTWKLKTQFNNEYSQDLDAINTKDKIGVELTGTBLADTLTLDSPIKVPLLLSEPINII 2010  
196 -----NVDSTI-----YF----- 204  
2011 DALEMRDAVEKPEFTIVAFVKYKQDVHSINLPPFTLQEVFFERNQTTIIVVENVOR 2070  
205 -----LG-----LTALLRYA----- 215

Db 2071 NLKHINIDQVRKYRAALGKLPOANDYLNSFNWQVSHAKEKLTALTKKYRITENDIQ 2130  
QY 216 ----- 215  
Db 2131 IALDDAKINFNEKLSQLQTMQIPDQYKOSYDLHDLKIAIANIIDBIIEIKLSLDBHYH 2190  
QY 216 ----- 215  
Db 2191 IRVNLVKTIHDLHLFIENIDFNKSGSSTASWQNVDTKYQIRIQIQEKLQQLKXHQINID 2250  
QY 216 -----QRNCTHSFVLNAMSRL-----FRVPKYING- 242  
Db 2251 IOHLAGLKGHIEAIDVRVLLDQLTGISFERINDVLEHVHKFVINLIGDFEVAEKINAF 2310  
QY 243 -----TKLKNWTKLKR--KQAPVKEQEFK----- 265  
Db 2311 RAKVHELIEREYVDQIQVLMKLVELTHQYKLETKIQUSNVLQQVKIKINDYPEKLVGFI 2370  
QY 266 ----- 265  
Db 2371 DDAVKLNLSPKFTFIEDVNKFLDMLIKLSFDYHQVDETDNKIREVTQRLNGEIQAL 2430  
QY 266 -----KAK----- 268  
Db 2431 ELPQKABALKFLEETKATVAVYLESQDRTKITLIINWLOEALSSASLAHMKAKPRETLE 2490  
QY 269 -----KTOSTTTPYFS-YTTSAAANVTNV-TYSITTAAARV 303  
Db 2491 DTRDRMYDMIOQELQYLSLVQVSTLVYISDWMTLAKNLDTFAEQYSIQDWAKRM 2550  
QY 304 STSTIAVRPDSF-----MKSIMATQ----- 324  
Db 2551 KALV-----EQGTVPEIKTILGTMPAFEVSLQALQKATQTPDFIVPLDRLRIPSVQIN 2605  
QY 325 ----- 324  
Db 2606 FKDLKNIKIPRSRSTPEFTILNTHIPSTFIDFVEMKVKIIRTIDQMNSLOMPVDIY 2665  
QY 325 LRDL----- 328  
Db 2666 LRDLKVEDIPLARITLPDFLPEIAPEFIPTPLNLDQVPLHPELPOLPHISHTIEV 2725  
QY 329 ----- 328  
Db 2726 PTFGLYSLIKIQSPFLTDANADIGNGTTSANEAGIAASITAKGSKLEVLNDFDQANA 2785  
QY 329 ----- 328  
Db 2786 QLSNPKNPLAKESVKFSSKYLRTEHSEMLPFGNAIEGKSNVVASLHTEKNTLELSNG 2845  
QY 329 -----ATWVYT 334  
Db 2846 VIVKINNQLTDSNTKYFHKLNIPKLPDSSQADLRNEIKTLLKAGHIAWTSKGSKWKA 2905  
QY 335 T-----LRYQON----- 341  
Db 2906 CPRFSDGTHESQISFTIEGPLTSFGLSNKINSKHLRVNQNLVYESGSLNFSKLEIQSOV 2965  
QY 342 -----PF-CBPSRNR 350  
Db 2966 DSQHVGHSVLTAKGMALFEGGKAEFTGRHDAHLNGKVIOTLKNLUFFSAPFEITASTNN 3025  
QY 351 -----TAVSEFMKN----- 359  
Db 3026 EGNLKVRFPLRLTGKIDFLNNVALFLSPAQQASQVARSFQYKYNQNFSAENNENIME 3085  
QY 360 THVLIRNET-----PYTYGT----- 375  
Db 3086 AHVGEINGEANLDFLNIPLETIPEMRLPYTIITPPLKDFSLWEKTLGKEPLKTKQSFDLS 3145  
QY 376 -----LWSSLYNET----- 386  
Db 3146 VRAQYKKNKHRHSITNPLAVLCEPISQISKSFDRHFEKRNALDPVTKSYNETKIKFKDK 3205

QY 387 ----- 386  
Db 3206 YKAKSHDELPRTFQIPGYTVPVNVVESPFTIEMSAFGYVPPKAVMSPFSILGSDVRV 3265  
QY 387 ----- 386  
Db 3266 PSVTLILPSLELPVLHVPRNLKLSLPHPKELCTISHIFIPAMGNITYDFPSKSSVITLNT 3325  
QY 387 -----MFVE--NK 392  
Db 3326 NAELFNQSDIVAHLLSSSSVIDALQYKLEGTTLTRKRGKLATALSLSNKFVEGSHNS 3385  
QY 393 TAS-----DSNKTTPT-SPSGMGFORFP----- 413  
Db 3386 TVSLTTKNNEVSAKTTKABEIPILRMNFQOELNGNTKSKPTVSSSMFEPKYDFNSSLMYST 3445  
QY 414 ----- 413  
Db 3446 AKGAVDHLKLSLESLTSYFSIESSTKGDVKGSVLSREYSGTIASEANTYLSKSTRSSVKL 3505  
QY 414 -----ID----- 415  
Db 3506 QGTSKIDDIWNLEVKENFAGEATLQRIYSLWEHSTKNHLQLEGLFPTNGEHTSKATLELS 3565  
QY 416 ----- 415  
Db 3566 PWOMGALVOVHASQSPSFHDPPDLGOEVALNANTKNQKIRWKNVRIIHSGSFQSOVELSN 3625  
QY 416 -----PLWDYL----- 421  
Db 3626 DQEKALHDIAGSLEGHRLFLKNIILPVVDKSLMDFLKLDDVTTSIGRRQHLRVSTAFVTK 3685  
QY 422 ----- 421  
Db 3686 NPNGYFSIPVKVLADKFTTGLKLNLSVLVMPFTFVPTDLOVPSCKLDFREBIQYK 3745  
QY 422 -----DSLLELDEI----- 430  
Db 3746 KLRTSSPALNPLTPEVKFPEVDVLTYSQPEDSLIPFEITVPESQLTVSFTLPKSVS 3805  
QY 431 ----- 430  
Db 3806 DGIAALDLNAVANKIADFELPTIIVPEQTEIIPSIFKSPAGIVIPSFQALTARPEVDSP 3865  
QY 431 -----R 431  
Db 3866 VYNATWSASLKNKADYVETVLDSTCSTVQFLEYELNVLTHKIEDGTLAGTKTGLAHR 3925  
QY 432 NFS-----LRSPT----- 439  
Db 3926 DFSAEYBEDGKEGLEWEGKAHLNKSAPFTDLHLRYQDKKKGISTSAASPAVGTGMD 3985  
QY 440 ----- 439  
Db 3986 MDEDDFSKWNFYSPQSSPDKLTI FKTCLRVEDSBETQIKVNWEEAASGLTSLKD 4045  
QY 440 -----YVN----- 442  
Db 4046 NVPKATGVLYDVNKYHWEHTGLTLREVSKLRRNLQNAEWVYQGAIRQIRIDIDVRFOK 4105  
QY 443 ----- 442  
Db 4106 AASGTTGYQEWKDKAQNLYQELLTQEQASFOGLKQNVFDGLVRVTQKFHKVKHLIDS 4165  
QY 443 ----- 442  
Db 4166 LIDFLNFRFOGPKGIYTTREELCTMFIREVGTVLQVSKVHNGSEILFSYFODLVIT 4225  
QY 443 ----- 442  
Db 4226 LPFELRKHLIDVISMVRELLKDLKSAQEVFKAIQSLKTTEVLRLNLODLLQFTQLIED 4285

QY 443 ----- 442  
Db 4286 NIKOLKEMKFTYLINVIQDEINTFNDYIPVFKLLKEN/CLNLHKNFNEFIONELQESQ 4345  
QY 443 ----- 442  
Db 4346 ELQIHQYIMALREYFDPISVGTWVKYVELEKIVSLIKNLIVALKDPHSEYIVSASNP 4405  
QY 443 -----LTPP----- 446  
Db 4406 TSQSSQVEQFLHRNIQEYLSILTDPDGKGEKIAELSAFAQBIKSAQATATKKIISDYH 4465  
QY 447 -----EHRANVLSN 458  
Db 4466 QPFRYKLQDPSDLSYDYKFAIESKRLIDLSQN 4500

RESULT 9  
US-08-750-717-2  
; Sequence 2, Application US/08750717  
; Patent No. 6180109  
; GENERAL INFORMATION:  
; APPLICANT: MOORMANN, Robertus J. M.  
; APPLICANT: VAN RIJN, Petrus A.  
; TITLE OF INVENTION: Nucleotide Sequences of Pestivirus  
; TITLE OF INVENTION: Strains, Polypeptides Encoded by These Sequences and Use  
; TITLE OF INVENTION: Thereof for Diagnosis and Prevention of Pestivirus  
; TITLE OF INVENTION: Infections  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: YOUNG & THOMPSON  
; STREET: 745 South 23rd Street  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/750,717  
; FILING DATE: 24-DEC-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 94201743.5  
; FILING DATE: 17-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/NL95/00214  
; FILING DATE: 16-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PATCH, Andrew J.  
; REGISTRATION NUMBER: 32,925  
; REFERENCE/DOCKET NUMBER: BO 39123  
; TELEPHONE: 703-521-2297  
; TELEFAX: 703-685-0573  
; TELEX: 248425 EMBON  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3898 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-750-717-2

Query Match 17.4%; Score 426; DB 3; Length 3898;  
Best Local Similarity 5.0%; Pred. No. 0.0059;  
Matches 191; Conservative 88; Mismatches 171; Indels 3373; Gaps 61;  
QY 1 MGRKEMW-----VRDVPK----- 13  
|||:|:

Db 18 MGVEEPVYDATGRPLPGDPSEVHPQSTLKLPHDRGRGNIKTTLKNLPRKDCRSNHLGP 77  
QY 14 -----MFVLISI 20  
Db 78 VSGIYVKPGPVFYQDYMGPVYHRAPLEFDEQCEVTKRIGRVTSQDGLYHYVICIDG 137  
QY 21 SFLLV-----SPINC----- 30  
Db 138 CILLKLAKGEPRTLKWIENFTDCLPWVTSQSDGASGSKKPKDRINKGKLIAPKEHE 137  
QY 31 -----KMSKALY----- 38  
Db 198 KDSRTRPPDATIVVEGVKYQVKKKVGKNTQDGLYHNKPPSRKLEKALLAWAVI 257  
QY 39 ----- 38  
Db 258 AIMLYQVVEAENITOWNLSDNGTNGIOHAMLYRGVNRSLHGIWPGBEICKGVPTHLATDVE 317  
QY 39 ----- 38  
Db 318 LIKEIQGMDASEGTNYTCCKLQREHWNKHGWCNWHNIDPWIOLMNRTOADLAEGPPVKEC 377  
QY 39 -----NRP----- 41  
Db 378 AVTCRYDKDADINNVTOQRNRPPTTLTGCKKGNFSPAGTVIESPCNFNVSVEBTLYGDEH 437  
QY 42 -----WRG----- 44  
Db 438 CGSLQLQDAALYLVDMGTMNTIENARQGAARVTSWLGRQLTAGKRLGRSKRTWFGAVALSP 497  
QY 45 -----L 45  
Db 498 YCNVTSKIGIWIYTNCTPACLPKNTKIIGPKGFTDPAEDGKILHEMGHLSFLLSLV 557  
QY 46 VLS-----KICKY-- 53  
Db 558 VLSDPAPETASALYLILHVIFQPHDEPEGCDTNQLNLTVELRTEDVIFSSVWNVGKYVC 617  
QY 54 -----KLDQLKLEILROLE----- 67  
Db 618 VRPDWMPYETEVALLFEEVGVVVKLALRALRDLTRVWNSASTIAFLICLIKVLRGQIVQG 677  
QY 68 -----TTISTKYN----- 75  
Db 678 VVWLLVLTGAOGLACKEDYRVAISSTDEIGLLGAGGLTTTWKEYNHDLQLNDGTVKASC 737  
QY 76 -----VSKQPVKNL----- 84  
Db 738 VAGSKVVTALNVSRKYLASLHKKALPTSVTFELLPGDGNPSTEEMGDDFRSGLCFPTS 797  
QY 85 -----TWNTE----- 89  
Db 798 PVVKGYNTLLNGSAFYLVCPIGTWGVIECTAVSPTTLRTEVWTKFRDRKPPPHRMDCV 857  
QY 90 -----FPQYVI----- 95  
Db 858 TTTVENEDLYCKLAGNWTVCVAGPVVYTGVLVKQRCWCGFDGPDGDLGPHYPIGRKILA 917  
QY 96 -----LAGP 99  
Db 918 NETGRIVDSTDCNRDGVVISTEGSHECLIGNTVKVKHASDERLGMPCRPKEIVSSAGP 977  
QY 100 IONYSITY-----LWFDYVSTQLRKP----- 120  
Db 978 VKTSTCTFNKTKLNRYVEPRDSYFQYMLKGEYQYWFEDLDTDRHSYPAEFVVLVVV 1037  
QY 121 -----AKYV----- 124  
Db 1038 ALLGGRYVLWLIYTVVLTQEAAGLPLQGEVVLIGNLIHTHDIIEVVVVFLLLYLDMRD 1097  
QY 125 -----YSQVNHAKTIT----- 136  
Db 1098 EPIKKWILLFHAMTNNPVKTTITVALLMVSVAKGKIDGGWQRLPGTSGFDIQLALTIV 1157  
|||:|:

QY 137 ----- 136  
Db 1158 VAVMLAKRDPPTVPLVITVAPLRTAKMTNGLSTDIAIAIATVSAALLTWTYISDYRYKWTW 1217  
QY 137 -----FRP----- 139  
Db 1218 LOYLIVTGTGIFLIRVLKGIGELDLHTPLPSHRPLFFILVYLITAVVTRWNLDIAGLL 1277  
QY 140 ----- 139  
Db 1278 LQCVPTLLMVTWADILTLILPTVELTKLYLKEVRIGAEGKWLWTKNFRVNDIYE 1337  
QY 140 ----- 139  
Db 1338 VDQAGEGYLPPSKQKTSMTGTMLPLIKALLISCVSNKNQFIYLLYLIFEVSYYLHKI 1397  
QY 140 ----- 139  
Db 1398 IDEIAGGTNFIISRLVAALIEVNWAFDNEEVRLGKLFLLSRVKELIHKHVRNEVMVRW 1457  
QY 140 -----PPCGRV-PSMTC----- 150  
Db 1458 FGDEEYVMPKLVGLVKAATLSKNKHCICTVCEDREWGETCPKCGRFGPPMTCGMTLA 1517  
QY 151 ----- 150  
Db 1518 DFEKHVKRIFFREDOSEGPVREEYAGLOYRARGQLFLRNLPLVATKVKQMLLVGNLGE 1577  
QY 151 -----LSEMLNV 157  
Db 1578 VGDLEHLGWLVRGPAVCKVTEHEKCTTSMMDKLTAFFGVMRGTTPRAPVRPTSLIKI 1637  
QY 158 SKRNDTG-----EOG 167  
Db 1638 RRLEGTWAYTHOGGISSVDHVTGCKDLLVCDTMRTRVVCQNNKMTDESEYGVKTDG 1697  
QY 168 C-----GNFT-----TFNPMFF-----NVPRWN----- 185  
Db 1698 CPEGARCYVFNPEAVNISGTGAMVHLQKTGGEFTCVTASGTAPAFDLKNLKMGSLPIP 1757  
QY 186 -----TKLYVG-----PTKVNDVDSOTI----- 202  
Db 1758 EASSGRVVRGVKGNEDSKPTKLMGSIQTVSKSTTDLTEMVKKITTSRGEFRQITLAT 1817  
QY 203 ----- 202  
Db 1818 GAKTTELPRSVIEEIGRHKRVLVLPLRAAESVYQYMRQKHPSTAFNLRIGEMKEGDM 1877  
QY 203 -----YFL-----GLTALLRYA-----QRNCT-----HSFVLNAMSR 231  
Db 1878 ATGITVASYGYFCQMPQPKLRAAMVEYSFIFLDEYHCATPEQLAIMGKIHRF-----SE 1931  
QY 232 NL-----FRVPKYINGTKLN----- 247  
Db 1932 NURVAMATPVGTVTGQKHPIEBFIAPDVMMKGDGLGSEYLDIAGLKIPIVEEMKSNML 1991  
QY 248 -----TMRKLKRK-----QAP----- 258  
Db 1992 VFVPTRNMAVEITAKLKARGYNSGYYSGEDPSNLRVVTSQSPYVVVATNAIESGVTLPD 2051  
QY 259 -----VKEQFEKKAK----- 268  
Db 2052 LDVVVDGLKCEKRIIRLSPMPPIVTLGRMAVTIGEQARGRVRGVKPGRYRSQETP 2111  
QY 269 ----- 268  
Db 2112 VGSKDHYDLOQORYIEDGINITKSPREMYNDWSLYEDSLMITQLBILNNLLISDEL 2171  
QY 269 -----KTQSTTTTPYFSYT----- 281  
Db 2172 PNAVKNMARTDHPPEIQLAYNSYEQVPLPKIKNGEVTDSYDNYTFLNARKLGDDVP 2231

QY 282 ----- 281  
Db 2232 PYYATEDEDLAVELLGLDWPDPGNOQTIVETGRALKQVVGSLSTAENALLVALLFGYVGYOA 2291  
QY 282 ----- 281  
Db 2292 LSKRHIPVVTDIYSIEDHRELDTHLQYAPNAIKTEGKETELKELAQGDVQRCVBEAMTWY 2351  
QY 282 -----TSAALNVTTNTY----- 294  
Db 2352 AREGIOFMKSQALKVKKETPTYKETMDTVTDYVVKFMEALADSKEDIKIKYGLMGTHTALYK 2411  
QY 295 -----SITTAARRVSTTIAY-----RPD-----S 314  
Db 2412 SISARLGGETAFATLVVVKWLAFGGESIADHVKQAATDLVVYIINRPQFPGETQODGR 2471  
QY 315 SFWKSIMATQLRDLATVWYTLRYR----- 339  
Db 2472 KEVASILLASA---LATYTKSNWYNNLSKIVEPALATLPYAATALKLPAPTRLESVIVLS 2528  
QY 340 -----QNP-----CEPSRNRITAVSE--- 355  
Db 2529 TAIYKTYISIRKGSDDLGTGVSAMEIMSQNPVSVGVIAVMLGVGCAVAAHNAIEASEOK 2588  
QY 356 -----FMKN-----THVLRNETP-----YTIYGTL----- 376  
Db 2589 RLLMKVFVKNFLDQAATDELVK-ESPEKIIMALFEAVQTVGNPLRVLVYHLGVFYKGWE 2647  
QY 377 -----DMSSLYNETMFMVENKTASDNKT--- 400  
Db 2648 AKELAQRTAGRNLFILIMFEAVELLGVDSEKVRQLSSNYILELLY---KPRDSIKSSV 2703  
QY 401 -----TPTSPSMGFQR----- 411  
Db 2704 REMAISWAPAPSCDWTPTDDRIGLUPQDNHFQVETKPCGYKMKAVKNCAGBELRLEBEG 2763  
QY 412 ----- 411  
Db 2764 SFLCRNKGSGSRNRYRVTKYDDNLLBIKPVIRMEGHVELYKYGATIKLDFFNNSKTILAT 2823  
QY 412 ----- 411  
Db 2824 DKWEVDHSTLVRVKRHTGAGYHGYLGEKPNHKLIERDCATITKDKVCFMKMRGCAF 2883  
QY 412 -----TFI----- 414  
Db 2884 TYDLSLNLTRLIELVHKNNLEDKEIPAATVTTWLAYTFVNEIDIGTIKPAFGEKVLEMQ 2943  
QY 415 ----- 414  
Db 2944 EBITLQPAVVVDTTDVAVTVVGEAPTMTTGTPTVFTSSGSLKQQVQLKLGVBGQYPG 3003  
QY 415 -----DPL----- 417  
Db 3004 TNPQASLHEAIOGADERPSVLILGSDKATSNRVKTAKNVKYVGRDPLEVRDMRGRKI 3063  
QY 418 -----WDYLDLSLLFLDEIRNFSLSRPTVNL----- 444  
Db 3064 LVVALSRVDNALLKFDYKGTFLTREALEALSGLRPPKKNITKABAQWLLCPEDQMEELP 3123  
QY 445 -----PP----- 444  
Db 3124 DWFAAGEPIFLEANIKHRYHLVGDIIATIKEKAKQLGATDGTIKIKEYGAKVYSMKLSNW 3183  
QY 445 -----PP----- 446  
Db 3184 VMQENKOGNLTPLFEELLQOCPPGQNKTAHMVSAYOLAQGNWMTSCHVFMGTVSARR 3243  
QY 447 -----EHR----- 449  
Db 3244 TKTHPYEAYVKLRELVEEHKMKTLCPGSSLGRHNDWIIGIKIKYQGNLRTKHLNLPKVAE 3303  
QY 450 ----- 449



Db 3304 QLCREGRHNVNKTSSVNTATGIRLEKLPVVRAQTDPNTHQAIRDKIDKEENLQTPG 3363  
QY 450 -----RAVN----- 453  
Db 3364 LHKLMVFNALKRPELESSYDAVEBELERGINRKGAGFFPERKNIGILDSEKNKVEE 3423  
QY 454 ----- 453  
Db 3424 IIDNLKGRNIKYETAIPIKNEKRDVNDWDAGDFVDEKKPRVIOYPEAKTRLAITKVMY 3483  
QY 454 ----- 453  
Db 3484 KWKQKPVVPIGVEGKTPFIQIPDKVKKWDQPNPAVSPDKAWDTQVTTKOLELIRD 3543  
QY 454 -----LSTNSL----- 460  
Db 3544 IQKYFKKKWHKFIDLTTHMSEVPVISADGEVYIRKGQSGQPDTSAGNSMLNVLTVW 3603  
QY 461 ----- 460  
Db 3604 YAFCEATGVPKSDFRVAKIHVCGDDGFLITERALGEKFAKGVQILYBAGKPQKITEGD 3663  
QY 461 ----- 460  
Db 3664 KMKVAYQDDIEFCSHTPIQVRWSDNTSSYMPGRNTTILAKWATRLDSSGERGTIAYEK 3723  
QY 461 -----W----- 461  
Db 3724 AVAPFLMYSWNPILIRICLLVLSTELQVKPKSTYYYEGDPISAYKEVIGHNLPDLK 3783  
QY 462 -----W-----WL 464  
Db 3784 RTSFEKLAKINLSMSVLGAWTRHTSKRLQDCVNIQVKGSGNL 3826

RESULT 10  
US-08-222-617A-2  
; Sequence 2, Application US/08222617A  
; Patent No. 5882879  
; GENERAL INFORMATION:  
; APPLICANT: Veenstra, Annemarie E.  
; APPLICANT: Martin, Juan F.  
; APPLICANT: Garcia, Bruno D.  
; APPLICANT: Gutierrez, Santiago  
; APPLICANT: Barredo, Jose L.  
; APPLICANT: Von Doehren, Hans  
; APPLICANT: Palissa, Harriet  
; APPLICANT: Van Liempt, Henk  
; APPLICANT: Montenegro, Eduardo P.  
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam  
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large  
; TITLE OF INVENTION: Quantities of ACV Synthetase  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/222,617A  
; FILING DATE: 04-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 97,157  
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3778 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-222-617A-2  
Query Match 17.2%; Score 422; DB 2; Length 3778;  
Best Local Similarity 5.2%; Pred. No. 0.0074;  
Matches 193; Conservative 69; Mismatches 171; Indels 3271; Gaps 55;  
QY 8 VRDVPK-----MFVLISISF----- 22  
Db 76 VNOVPERCDLSGLTDTSTRYQLASTGFGDASAAQERLMTVPVDVHAALQELCLERRSV 135  
QY 23 ----- 22  
Db 136 GSVINFSVHQLKGFNGTHTTITASLHREQLNQLNSPSPVSVPTIVTHENRDGWSVAQAV 195  
QY 23 -----LVSFIN-----CKV- 32  
Db 196 ESIEAARGSEKESVTAIDSASSLVKMLFDLLVSVFVDDARIPCFDFPLAVIVRECDAN 255  
QY 33 -----MSKALYN-----RWRG-- 44  
Db 256 LSLTLRFSDCLFNEETICNFTDALMILLAEAVIGRVTVPVADIELLSAEKQOLEWNTD 315  
QY 45 ----- 44  
Db 316 GEYPSKRLHLIEEVRHEDKIAVVCDERELTYGELNAQNSLARYLRSIGILPEQLV 375  
QY 45 -----LVLSKIGKYK-----LDQLK----- 59  
Db 376 ALFLDKSEKLIIVTILGVNKGSAAYVIDPTYDPRVFLVDDTKARAIIASNQHVRLQR 435  
QY 60 -----LEILRQLETTISTKYNVSKQPVKNL-----TMNTEFQO----- 92  
Db 436 EVIGDRNLICIRLEPLLASLAQSSKFFPAHNLDDPLTSQQLAYVYVTSGTGTFPGKPIK 495  
QY 93 -----YVILAGP----- 99  
Db 496 QHTNVVNSITDLSARYGVAGQHHAILLFSACVFPFVROTLMALVNGHLLAVINDVEKY 555  
QY 100 -----IQNVSYTYL-----WDFP----- 112  
Db 556 DADTLPLPFIIRHSITYLNGTASVLOEYDFDCPSLNRILVGENLTEARYLALRQRFKNR 615  
QY 113 -----YSTQLRKPAK-----YVYSQYNHTAKTITFRPPPCG 143  
Db 616 ILNEYGFTESAFVTALKIFDPESTRKDTSLGRPVNRNVKCYILN-----PSLK 662  
QY 144 RVPSTMTCLSEM-----LNVSK----- 159  
Db 663 RVP-IGATGELHIGGLGIGISKGYLNPPELTPHPIPNPQTDCQCEKQGLNSLMYKTDGLAR 721  
QY 160 -----RNDTGEQCGNFTTFNPFNFVPRWNTKL----- 188  
Db 722 WLPNGEVEYLGRADFQIKLGRIRIFGE-----IETMLAMYPRVTSLVVSKKLNRG 773  
QY 189 -----YVGPTK-----VNYD----- 198  
Db 774 PEETTNEHLVGYVCDVSASVSADLLSFLEKLPYIMPTIRLVQLSQIPVNVNGKADLRA 833  
QY 199 -----SOTYVFLG----- 206  
Db 834 LPVDISNSTEVRSLRGDTETALGEIWDVLGARQSVSRNDNFRGLGHSITCIQLIA 893  
QY 207 -----LTALLRYAQRNC-----THSFYL 225  
Db 894 RIRQRQLSVSISVEDVFATRTLERWADLLQNKQCKDKPHEAPTELLEENATDNIYL 953  
QY 226 VNAMSR----- 231

Db 954 ANSLQOQGVVHYLKSMEQSDAYVMQSVLRYNTTSLPDLFORAMKHAQQSPALRLRFSWE 1013  
Qy 232 -----NLFRV----- 236  
Db 1014 KEVQOLLDDQPPDLWRFLYFTDVAAGAVEDRKLEDLRRQDLTERFKLDVGRLFVRVLIKH 1073  
Qy 237 ----- 236  
Db 1074 SENRFTCLFSCHAILDWSLPLLFKEKVHETYLQLLHGDNLTSMDPPYTRTQYLHAHR 1133  
Qy 237 -----PKY----- 239  
Db 1134 EDHLDFWAGVVQKINERCDMALLNERSRYKVLADYDQVEQRHVHTIALSGDAWLADLR 1193  
Qy 240 -----INGT----- 243  
Db 1194 QTCSAQGITLHSLQFVWHAVLHAYGGTHYITGTTISGRNLPILGIERAVGPIYNTLPL 1253  
Qy 244 -----KLKNTWRKLRKQAPVK----- 260  
Db 1254 VLDHSTPKDKTIMEATEDVQAKVMNVRSGNVELGRHLKTDLKHGLFDSLVLNYPNLD 1313  
Qy 261 ----- 260  
Db 1314 KSRTLEHQTELGYIEGGTEKLNPLAVIAREVETGGFTVSYCYASELPEEVMISELLH 1373  
Qy 261 -----EOFEKKAKK----- 269  
Db 1374 MVQDTLMQVARGLNEPVGSLEYLSSIQLEQAAWNATEAFPPDTTILHEMFENEASQPKDK 1433  
Qy 270 ----- 269  
Db 1434 IAVVYBETSITYRELNERANMAHQLRSDVSPNPNVIALVMDKSEHMTIVILAVKSGG 1493  
Qy 270 ----- 269  
Db 1494 AYVPIDPGPNDRIOVILEDTQALAVIADSCYLPRIKGMASGTLLYPSVLPANPDSKWS 1553  
Qy 270 -----TOSTTTPYFSYT----- 281  
Db 1554 VSNPSPLSRSTDLAYIYTSGTGRPKGVTVHGGVNVNQLVSLKVPGLRDTDDDEVILSP 1613  
Qy 282 -----TSAALNVT-----NVTY----- 294  
Db 1614 SNYVDFHVEQMTDALNGQTLVLNDGMGRDKERLYRIEKNRVTYLSGTFSVSMYEF 1673  
Qy 295 -----SITT----- 298  
Db 1674 SRFKDHLRRVDCVGEAFSEPVDPKIRETFHGLVINGYGTEVSITTHKRLYPFERMDK 1733  
Qy 299 ----- 298  
Db 1734 SIGQQVHNSTSYVLNEDMKRTPIGAVGELYLGEGVVGVGNRADVTAERFIPNPPQSEE 1793  
Qy 299 ----- 298  
Db 1794 DKREGNSRLYKTDLVRWIPGSGSEVYLGGRNDFQVKIRGLRIEVEGTEIAYLSSYHGK 1853  
Qy 299 -----AARR----- 302  
Db 1854 QSVVIACDCEGAQKFLVGYVADAALPSAAIRRFMQSRPLPGMVPSRLILVSKFPVTPS 1913  
Qy 303 ----- 302  
Db 1914 GKLDTKALPPABESEIDVVPFRSEIERSLCDIWAELLEMHPPEIGIYSDPFSLGDSLK 1973  
Qy 303 -----VSTSTI----- 308  
Db 1974 STKLSFMIHESFNRAVSALFCHRTVEAQTHLILNDAADVHEITPIDCNDTQMIPVSR 2033  
Qy 309 -----AYRDPDSGF----- 316  
Db 2034 QERLLFIHFENGSNAYNIDAFELPGSVDSILLEGALRGNLARHEALRTLVLKDHATGI 2093

Qy 317 ----- 316  
Db 2094 YLQKVLSPDEAQGMFSVNVDTAKQVERLDQETASLSQHVFRLLDDELWPWEARILKLESGGL 2153  
Qy 317 ----- 316  
Db 2154 YLILAFPHHTCFDAWSLKVFEQELRALYAAALQTKSAANLPALKQAQYKEYALVHRRQLSGD 2213  
Qy 317 ----- 316  
Db 2214 RMRNLSDFWLRLIGLEPLQLITDRPPVQPKYDGDGLSIELSKKETENLRGVAKRCKSS 2273  
Qy 317 ----- 316  
Db 2274 LYVVLVSVCVMLASYANQSDSVSGIPVSHRTHPOFQSVIGFVNLVLRVDISQSAICG 2333  
Qy 317 -----MKSIMATQLRD----- 327  
Db 2334 LIRRVMKELVDAQLHQDMPFQEBVTKLQVDNDPSRHLVQNVFNFESRANGEDHARSEDE 2393  
Qy 328 ----- 327  
Db 2394 GSLAFNQVRPQVPDVSVAKFDLNATVTELESGLRVNFNYATSLFNKSTIQGFLHTYEYLL 2453  
Qy 328 ----- 327  
Db 2454 RQSELSEAGINEDTQLSLVRPTENGDLHLPLAQSPLATTAEEQKVASLNQAFEREAFLA 2513  
Qy 328 -----LATW- 331  
Db 2514 AEKIAVQGDRLSYADLNGQANQLARYTQSVSCIGADDGIALMLEKSIDTIIICILAIWK 2573  
Qy 332 ----- 331  
Db 2574 AGAAVPLDPTYPGRVOLILEEIKAKAVLVHSSHASKCERHAKVIAVDSPAIETAVSQ 2633  
Qy 332 ----- 331  
Db 2634 QSAADLPTIASIGNLAYIIFTSGTSKPKGVLVBEQKAVLLLRDALRERYFGDRCTKHGV 2693  
Qy 332 ----- 331  
Db 2694 LFLSNYVDFSVQVLVLSVSHGKLIIVPPAEFVADDEBFYRMASHTGLSYLSGTPSLLOKI 2753  
Qy 332 -----VY----- 333  
Db 2754 DLARLDHLQVTAAGEELHATQEKORRRFNGPIYNAVGVTTVTYVNIIAEFTTNSIFEN 2813  
Qy 334 -----TTLRYRQNPFCBP- 346  
Db 2814 ALREVLPOTRAVVLTAALQPPVFDVAGELYLAGDVTVRGYLNQPLLTQORFIPNPFCKEE 2873  
Qy 347 -----SR-----NRTAVSE----- 355  
Db 2874 DIAMGRFARLYKTGDLVRSRNRQOQPOLEYLGRGDLQIKMRGYRIEISEVQNVLTSPPG 2933  
Qy 356 -----FMKNTHVLI-----RNET----- 368  
Db 2934 VREGNAVAKYENDTYSRTAHSLVGYYYTDTNETVSEADILTFMKARLTYMVPVSHLCLE 2993  
Qy 369 -----PYTIYGLDMSSL----- 381  
Db 2994 GALPVTINGKLDVRRLPEIINDSAQSSYSPPNRIIEAKMCRLESALGMRGCGIDDDLFK 3053  
Qy 382 ----- 381  
Db 3054 LGDSITSILHLVAQIHQVGCCKITVRDIFEHRTARALHDHVMKDSRDSNVQTRTEQGP 3113  
Qy 382 -----YNETMFV----- 389  
Db 3114 VIGEAPLLPIQDWFLSKALQHPMTWNHTFYVTRPELVDVLSAAVRDLQOYHDFVFMRLK 3173

390 ----- 389  
3174 REEVGVQSAEDFSPAQLRVNLVKDVGSAANEILDGSGFNGLENGPIGSIYLHG 3233  
390 ----- ENKTASDSN 398  
3234 EDRSARVWFVHHMAIDTVSQIILVRDLQTLRYRNGSLGSGSSPRQWAEAIQNTKASDSE 3293  
399 KT ----- 400  
3294 RNHWKLVMTASSISALPTSTGSRVRLSRSLSPKTSALIQGGIDRQDVSVYDSLTSV 3353  
401 ----- TPTSPM----- GQRTFID----- 415  
3354 GLAQHIAPTGPMSVMTIEGHGREVDQTLDSRTMGWFTTMYPEIPRLSTENIVQGWA 3413  
416 ----- 415  
3414 VSRFPQVARGVGYGLYGTQHPQLQVTVNYLGLQARKQSKPEWLVAGONEPEYGL 3473  
416 ----- 415  
3474 MTPEDKDRSSAVDTAVCIDGTMIIIDVDSANSLSESEQFISSIEGLNKILDGRASQ 3533  
416 ----- PLWDYLD----- 422  
3534 TSRPPDPQPAETVTPYFEYLEPPRQGTFLFLPPGEGGAESYFNNIVKRLQTNMVFN 3593  
423 ----- SLLFDEIRNFSLRSP----- 438  
3594 NYLHSLRLRTFELAEYLDQVRGIQPHGPHYHFGISGILAMEMRSLVASDEKIGF 3653  
439 ----- TYVN----- LTPPEH----- 448  
3654 LGIIDTFVNRGATRTIGLGTGLDPIHHIYNPDNANFORLPSATDRIVLFKMRPNK 3713  
449 ----- RR----- AVNLSTSNLW 462  
3714 YESENQRRLVEYDRTLRGLDLSLLPSDSVQLVPLTDTHFSW 3757

RESULT 11  
US-08-222-617A-12  
; Sequence 12, Application US/08222617A  
; Patent No. 5883879  
; GENERAL INFORMATION:  
; APPLICANT: Veenstra, Annemarie E.  
; APPLICANT: Martin, Juan F.  
; APPLICANT: Garcia, Bruno D.  
; APPLICANT: Gutierrez, Santiago  
; APPLICANT: Barredo, Jose L.  
; APPLICANT: Von Doehren, Hans  
; APPLICANT: Palissa, Harriet  
; APPLICANT: Van Liempt, Henk  
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam  
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large  
; TITLE OF INVENTION: Quantities of ACV Synthetase  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/222.617A

FILING DATE: 04-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
REFERENCE/DOCKET NUMBER: 97,157  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3666 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Penicillium chrysogenum  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..3666  
OTHER INFORMATION: /label= region  
OTHER INFORMATION: /note= "Region of ACV Synthetase; aa 62-372"  
US-08-222-617A-12  
Query Match 17.1%; Score 420; DB 2; Length 3666;  
Best Local Similarity 5.2%; Pred. No. 0.008;  
Matches 188; Conservative 68; Mismatches 163; Indels 3211; Gaps 53;  
QY 8 VRDVPK-----MEVLISISF----- 22  
DB 15 VNQVPERCDLSGLTDTSTRYQLASTGFGDASAAQERLMTVPVDVHAALQELCLERRSV 74  
QY 23 ----- 22  
DB 75 GSVINFSVHOMLKGFGNGTHTITASLHREONLQNSPSWVSPVTVTHENRDGWSVAQV 134  
QY 23 -----LLVSPIN-----CKV- 32  
DB 135 ESTEAAARGSEKSVTAIDSAASLVKMGFLFDLLVSFVDADARIPCFDFPLAVIVRECDAN 194  
QY 33 -----MSKALYN-----RPMRG-- 44  
DB 195 LSLTLRPSDCLFNEETICNFTDALNILLABAVIGRTVPVADIELLSAEQKQLEWNTD 254  
QY 45 -----LVLSKIGKYK-----LDOLK----- 44  
DB 255 GEYPSKRLHLLIEVVERHEDIKAVVCDERELTYGELNAGNSLARYLSIGILPEQLV 314  
QY 45 -----LVLSKIGKYK-----LDOLK----- 59  
DB 315 ALFLDKSEKLIIVTLGVKWSGAAVPIIDPTYPDERVRFVLDLTKARAIASNOHVERLQR 374  
QY 60 -----LEILRQLETTISTKYNVSKQPVKNL-----TWNTBPPQ----- 92  
DB 375 EVIGDRNLCTIIRLEPLLASLAQSSKFFPAHNLDDLPLTSQOLAYVTVTSGTTFPKGIFK 434  
QY 93 -----YYILAGP----- 99  
DB 435 QHTNVVNSITLSARYGVAGQHEAILLFSACVFEPVROTLMAVNGHLLAVINDVEKY 494  
QY 100 -----TONYSITYL-----WFDP----- 112  
DB 495 DADTLPLFFIRRHISITYLNGTASVLYQYDFSDCPSLNRIILVNGENLSEARLYALRQRPKNR 554  
QY 113 -----YSTOLRKPAAK-----YVSYQYNHTAKTITFRPPPCG 143  
DB 555 ILNEYGFTSAFTALKIFDPSTRKDTSLGRPVNVKCYILN-----PSLK 601  
QY 144 RPSMTCLSEM-----LNVSK----- 159  
DB 602 RVP-IGATGELHIGGLGISKGLNLRPELTPIPNPFDTCCKQLGINSIMYKTDGLAR 660  
QY 160 -----RNDTGEQCCGNFTTNPMFFNVPRWNTKL----- 188  
DB 661 WLPNGEVEYLGRADFOIKLRGIRIEFG-----IETMLAMYPRVRTLVSWSKKLRNG 712

QY 189 -----YVGPTK-----VNVD----- 198  
Db 713 PBETTNEHLGVYVCDASVSEADLLSFLKELPRYMPITRLVQLSQIPVNVNGKADLRA 772  
QY 199 -----SQTYYFLG----- 206  
Db 773 LPAVDISNTEVRSLRGDTGTEIALGEIWAIDLGARQORSVRNDNPFRLGGHSITCIQILIA 832  
QY 207 -----LTALLRYAQRNC-----THSFYL 225  
Db 833 RIRQORLSVSISVEDVFATRTLERMADLLQNKQOQKCDKPHEAPTELLEENAAATDNIYL 892  
QY 226 VNAMSR----- 231  
Db 893 ANSLQOGFVYHYLKSMEQSDAYVMSQSVLRYNTTILSPDLFORAKHQAQSPFALRLRFSWE 952  
QY 232 -----NLPRV----- 236  
Db 953 KEVFQLLDQDPLDWRFLYFTDVAAGAVEDRKLDRQDLTERFKLDVGRLFRVYLKIH 1012  
QY 237 ----- 236  
Db 1013 SENRFTCLFSCHAALDGSPLPLFEKVHETYLQLLHGDNLTSSMDDPYTRTORYLHAHR 1072  
QY 237 -----PKY----- 239  
Db 1073 EDHLDWAGVVQKINERCDMALLNERSRYKVQLADYDQEQORHVITIALSGDAWLADLR 1132  
QY 240 -----INGT----- 243  
Db 1133 QTCSAQGITLHSILOFVWHAHVHAYGGGTHITITGTSRNLPLGIERAVGPYINTLPL 1192  
QY 244 -----KLNTMRKLKRKQAPVK-----EQUEKAKK----- 269  
Db 1193 VLDHSTFKDIMEAIEDVQAKVNMNSRGVNLGRHLKTDLKHGLFDSLFLVLENYPNLD 1252  
QY 261 ----- 260  
Db 1253 KSRTLEHQTELGYIEGGTEKLNPLAVIAREVETGGFTVSICYASELPEEVMISELLH 1312  
QY 261 -----EQUEKAKK----- 269  
Db 1313 MVQDTLMQVARGLNEPVGSLLEYLSSIOLEQAAWNATEAFPDITLHEMFENEASQKPK 1372  
QY 270 ----- 269  
Db 1373 IAVVYEETSLTYRELNERANMAHQLRSDVSPNPNEVIALVMDKSEHMIWNILAVKWSGG 1432  
QY 270 ----- 269  
Db 1433 AYVPIDPGYPNDRIQIYILEDQALAVIADSCYLPRIKMAASGTLVPSVLPANPDSKWS 1492  
QY 270 -----TQSTTTTPYPSYT----- 281  
Db 1493 VSNPSPLSRSTDIAYIITYSGTGRPKGVTVHEHHGVNVLQVLSKVFGRLDDEVLFSF 1552  
QY 282 -----TSAALNVTT-----NVTY----- 294  
Db 1553 SNVYFDHFVQMTDAILNGTLLVNDGMGRDKERLYRIEKNRVITYLSGTPSVWSMYEP 1612  
QY 295 -----SITT----- 298  
Db 1613 SRFKDHLRRVDCGEAFSEVPFKIRETFHGLVINGYGPTEVSITTHKRLLPFPERRMDK 1672  
QY 299 ----- 298  
Db 1673 SIGQQVHNSTSYVLNEDMKETPIGAVGELYLGEGVVRGYNRADVTABRFIPNPFQSEE 1732  
QY 299 ----- 298  
Db 1733 DKREGRNRLYKTGDLVRWIPGSSGEVEYLGRNDFQVKIRGLRIELGETEAILSSVHGK 1792

QY 299 -----AABR----- 302  
Db 1793 QSVVIAKDCREGAQKFLNGVYVADAALPSAATIRRFMQSRPLPGYVWPSRLILVSKFPVTPS 1852  
QY 303 ----- 302  
Db 1853 GKLDTKALPABESESIDVVPSEIERSLCDIWAELELMEHPPEIGIYSDFFSLGGDSLK 1912  
QY 303 -----VSTSTI----- 308  
Db 1913 STKLSFMIHESFNRAVSVALFCHRTVEAQTHLILDAADVHEITPIDCNDTQWIPVURA 1972  
QY 309 -----AYRPSDF----- 316  
Db 1973 QERLLFIHFENGSNAYNIDAAFELPGSDVASLLEQALRGNLARHDEALRTLVLKDHATGI 2032  
QY 317 ----- 316  
Db 2033 YLQKVLSPDEAQMFSVNVDTAKQVERLQDEIASLSQHVPRLLDDELPEWARILKLESGL 2092  
QY 317 ----- 316  
Db 2093 YLILAFHHTCFDAWSLKVFEQELRALYAALOKTKSAANLPALKAQYKEYALYHRQLSGD 2152  
QY 317 ----- 316  
Db 2153 RMENLSDFWLRLKLI GLEPLQLITDRPRVQPKYDGGDDLIELSKKETENLRGVAKRCKSS 2212  
QY 317 ----- 316  
Db 2213 LYVVLVSVYCVMLASYANQSDVSGIPVSHRTHPQFSQVIGPFVNLVLRVDISQSAICG 2272  
QY 317 -----MKSIMATQLRD----- 327  
Db 2273 LIRRVNKLVDQLQHDMPFQSVTKLQVDNDPSRHPLVQNVNPFESRANGHARSEDE 2332  
QY 328 ----- 327  
Db 2333 GSIAFNQYRPQVDSVAKFDLNAIVTELESGLRVNFNATSLFNKSTIOGFLHTYEYLL 2392  
QY 328 ----- 327  
Db 2393 RQLSLSABGINEDTQLSILVRPTENGDLHLPLAQSPPLATTABEQKVASINQAFEREAFLA 2452  
QY 328 -----LATW----- 331  
Db 2453 AEKIAVQGDRLSYADLNGQANQLARYIOSVSGIGADDGIALMLEKSIDTITCILATWK 2512  
QY 332 ----- 331  
Db 2513 AGAAVPLDPTYPGRVOLILBEEIKAKAVLVHSSHASKERHGAKVIAVDSPAIETAVSQ 2572  
QY 332 ----- 331  
Db 2573 QSAADLPTIASIGNLAYIIFTSGTGKPGVLVEQKAVLLLRDLALRERYFRGDCCTKHGV 2632  
QY 332 ----- 331  
Db 2633 LFLSNVYVDFSVQQLVLSVSGHKLIVPPAEFVADDEFYRMASHTGLSYLSGTPSLLQKI 2692  
QY 332 -----VY----- 333  
Db 2693 DLARLDHLQVVVTAAGEELHATQYEKMRFRFNGPIYNAVGVTTVTYVNIIEAFTTNSIFEN 2752  
QY 334 -----TTLRYRONPFCEP----- 346  
Db 2753 ALREVLPOTRAYVNAALQPVFPAVGELYLAGDVTVRGYNLQPLLDQRFIPNPFCKEE 2812  
QY 347 -----SR-----NRTAVSE----- 355  
Db 2813 DIAMGRFARLYKTGDLVRSRFRNQOQPLEVLGRGDLQIKMRGYRIEISEVQNVLTSSPG 2872  
QY 356 -----FMKNTHVLI-----RNET----- 368

Db 2873 VREGAVVAKYENNDTSRTAHSVGVYTTDNETVSEADILTFMKARLPTVMVPSHLCCLE 2932  
QY 369 ---PYTIYGLDMSSL---  
Db 2933 GALPVTINGKLDVRLPEINDSAQSSYSPRNIEAKMCLWESALGHERCIGIDDDLFK 2992  
QY 382 ---  
Db 2993 LGDSITSLHLVAQHNNQCKITVRDIPEHRTARALHDHVFMKDSRNVTOFRTEQGP 3052  
QY 382 ---YNETMPV---  
Db 3053 VIGEAPLLPIQDWFLSKALQHPMYNHTFYVTRPELDVDSLSAAVRDLOOYHDVFRMLK 3112  
QY 390 ---  
Db 3113 REBVGVSQFAEDFSPAQLRVLVNKVDGSAANVEILDGWSQFNLENGPIGSIGYLHG 3172  
QY 390 ---ENKTASDSN 398  
Db 3173 EDRSARVWFSVHMAIDTVSQILVRDLQTLVRNGSLGSKGSSFRWAEAIQNYKASDSE 3232  
QY 399 KT-----  
Db 3233 RNHNKLVMETASSISALPTSTGSRVRLSRSLSPKTSALIOGGIDRDQDVSVVYDSLLTSV 3292  
QY 401 ---TPTSPM---GFQRTFD---  
Db 3293 GLALQHIATGSPMVTIEGHEVDQTLDSRTMGWFTMYPPPEIPRLSTENIVQGVVA 3352  
QY 416 ---  
Db 3353 VSRFRQVPARGVGYGLYGTQHPVQTVNYLVQLARKQSKPKWVLAVGDNPEYGL 3412  
QY 416 ---  
Db 3413 MTSPEKDRSSAVOVTAVIDCTMIIDVDSAWSLESEQFISSEBGLNKILDGRASQ 3472  
QY 416 ---PLWDYLD---  
Db 3473 TSPFPDPQPAETTYTYFYLEPPROGPTLFLPPGEGGAESYFNIVKRLQRTMNVFN 3532  
QY 423 ---SLLFDEIRNFSRSP---  
Db 3533 NYVLHSLKRLTPBELAEMVLDQVRGIQPHGPHYFICWSFGILAMEMSRRLVASDEKIGF 3592  
QY 439 ---TYVN-----LTPPEH 448  
Db 3593 LGIIDTYFNVRGATRTIGLDTILDPIIH 3622

RESULT 12  
US-08-222-617A-27  
; Sequence 27, Application US/08222617A  
; Patent No. 5882879  
; GENERAL INFORMATION:  
; APPLICANT: Veenstra, Annemarie E.  
; APPLICANT: Martin, Juan P.  
; APPLICANT: Garcia, Bruno D.  
; APPLICANT: Gutierrez, Santiago  
; APPLICANT: Barredo, Jose L.  
; APPLICANT: Von Doehren, Hans  
; APPLICANT: Palisea, Harriet  
; APPLICANT: Van Liempt, Henk  
; APPLICANT: Montenegro, Eduardo P.  
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam  
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large  
; TITLE OF INVENTION: Quantities of ACV Synthetase  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive

; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/222,617A  
; FILING DATE: 04-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 97,157  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3727 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-222-617A-27

Query Match 17.1%; Score 420; DB 2; Length 3727;  
Best Local Similarity 5.2%; Pred.No. 0.0084;  
Matches 188; Conservative 68; Mismatches 163; Indels 3211; Gaps 53;  
QY 8 VRDVPK-----MFVLISISF----- 22  
Db 76 VNOVPERCDLSGLTDTSTRYQLASTGFGDASAAQYERLATVPVDVHAALQELCLERRSV 135  
QY 23 ----- 22  
Db 136 GSVINFSVHQLKFGNGTHTTITASHRQNLQNSPSSVWVSPTIVTHENRDGWSVAQAV 195  
QY 23 -----LLVSFIN-----CKV- 32  
Db 196 ESIEAARGSEKSVTAIDSASSLVKMGFLPDLLVSFVDADARIPCFDFPLAVIVRECDAN 255  
QY 33 -----MSKALYN-----RPMRG-- 44  
Db 256 LSLTRFSDCLFNEETICNFTDALNILLAEAVIGRVTVPVADIELLSAEQKQLEENNTD 315  
QY 45 ----- 44  
Db 316 GEYPSKRLHLIEEWEHERHEDKIAVVCDERELTYGELNAQNSLARYLRSIGILPEQLV 375  
QY 45 -----LVLSKIGKYK-----LDQLK----- 59  
Db 376 ALFLDKSEKLIIVTILGVKWSGAAYVPIDPTYDPERVRFVLDLDDTKARAIIASNOHVERLQ 435  
QY 60 -----LEILRQLETTISTKYNVSKQPVKNL-----TWNTFPO--- 92  
Db 436 EVIGDRNLCIIRLEPLLASLAQDSSKFPANLDDLPLTSQQLAYVYTTSGTTCFPKGIK 495  
QY 93 -----YYILAGP----- 99  
Db 496 QHTNVVNSITDLSARYGVAGQHEAILLPSACVPEFVRQTLMALVNGHLLAVINDVEKY 555  
QY 100 -----IQNYSIYTL-----WPDF----- 112  
Db 556 DADTLLPFIIRHSITYLNGTASVLOQYDFSDCPSLNRILVGENLFEARYLALRQPFKNR 615  
QY 113 -----YSTQLRKPAC-----YVSQYNHTAKTITFRPPPCG 143  
Db 616 ILNEVGFTESAFTALKIFDPPESTRKDTSLGRPVNRVKCYILN-----PSLK 662  
QY 144 RYPSMTCLESM-----LNVSK----- 159  
Db 663 RVP-IGATGELHIGGIGSKGYLNRPELTPHRIPNPFQDCEKQIGNSLMVKTGDLAR 721  
QY 160 -----RNDTGEQCGNFTTFNPFNVRNWKTL----- 188

```
Db 722 WLPNGEVEYLGRADFOIKLRGIRIEGE-----IETMLAMPVVRTSLVSKLRNG 773
QY 189 -----YVGPTK-----VNV----- 198
Db 774 PEETTNEHLVGYVCDASVSADLLSFLKELPRYMIPTRLVQLSQIPVNVGKADURA 833
QY 199 -----SQTIFYFLG----- 206
Db 834 LPAVDISNTEVRSLRGDTETALGEIWADVLGARQSRVSRNDNFRGLGHSITCICLIA 893
QY 207 -----LTALLRYAQRNC-----THSFYL 225
Db 894 RIRQRQLSVSISVEDVFATRILERMADLLQNKQCKDPHEAPTELEBENAATDNIYL 953
QY 226 VNAMSR----- 231
Db 954 ANSLQGGFYHYLKSMEQSDAYVMSQVLRNTLTSPDLFORANKHAQSPPALRLRSWE 1013
QY 232 -----NLFRV----- 236
Db 1014 KEVFOLLQDPPDLWRFLYFTDVAAGAVEDRKLERQDLTERFKLDVGRLFRVYLKH 1073
QY 237 ----- 236
Db 1074 SENRFTCLFSCHAILDCGWSLPLFEKVHETYLQLLHGDNLTSMDDDPYTRTQRYLHAHR 1133
QY 237 -----PKY----- 239
Db 1134 EDHLDFAWGVQKINERCDMALLNERSRYKVQLADYDQEQORHVITIALSGDAWLADR 1193
QY 240 -----INGT----- 243
Db 1194 QCSAQGITLHSILOFWHVAHVHAYGGTHTITGTTISGRNLPILGIERAVGPYINTLPL 1253
QY 244 -----KLKNTMRKLRKQAPVK----- 260
Db 1254 VLDHSTFKDKTIMEAIEDVQAKVNMVNRGNVGLRLKHTDLKHGLFDSLVLNYPNLD 1313
QY 261 ----- 260
Db 1314 KSRTLEHQTELGYSIEGGTEKLNYP LAVIAREVETGGFTVSYCIYASELFEYMISELLH 1373
QY 261 -----EOFEKKAKK----- 269
Db 1374 MVQDTLMQVARGLNEPVGSLLEYLSSIOLEQAAWNATEAFPDTHHEMFENERSOKPK 1433
QY 270 ----- 269
Db 1434 IAVVYBETSLTYRELNERANMAHQLRSDVSPNPNVIALVMDKSBHMIVNILAVWKSOG 1493
QY 270 ----- 269
Db 1494 AYVPIDPGYNDRIQVILETOQALAVIADSCYLPRIKGMASGTLLYPSVLPANPDSKWS 1553
QY 270 -----TQSTTTPYFSYT----- 281
Db 1554 VSNPSPLSRSTDLAYIIYTGTTGRPKGVTVHEHGVNVLQVSLSKYVGLRDTDDDEVILSF 1613
QY 282 -----TSAALNVTT-----NVTY----- 294
Db 1614 SNTVPDHFVEQMTDAILNGQTLVLVNDGMGRDKERLYRIEKNRVITGLSTPVSVMWEF 1673
QY 295 -----SITT----- 298
Db 1674 SRFKDHLRRVDCVGEAFSPVPDKIRETFHGLVINGYGTEVSITTHKRLYFPERRMDK 1733
QY 299 ----- 298
Db 1734 SIGQQVHNSTSVYLNEDMKRTPIGAVGELYLGEGVVRGYNRADVTAERFIPNPFQSEE 1793
QY 299 ----- 298
Db 1794 DKREGNSRLYKTGDLVRWIPGSSGEVYLRNDFOVKIRGURIENVGEIATILSSYHGIK 1853
```

---

```
QY 299 -----AARR----- 302
Db 1854 QSVVIACDCREGAQKFLVGYVADAAALPSAAIRRFMQSRLPGYMVP SRLILVSKFPWTPS 1913
QY 303 ----- 302
Db 1914 GKLDTKALPABESESIDVPPRSEIERSLCDIWASLELLEMPHEEIGIYSDOFFSLGDSLK 1973
QY 303 -----VSTSTI----- 308
Db 1974 STKLSFMIHESFNRAVSVALFCHRTVBEAQTHLILNDAADVHEITPIDCNDTQMI PVSRA 2033
QY 309 -----AYRPDSSP----- 316
Db 2034 QERLLFIHFENGSNAYNIDAAPELPFGSVDAISLEQALRGNLARHEALRTLVLKDHATGI 2093
QY 317 ----- 316
Db 2094 YLQKVLSPDEAQGMFSVNVDTAKQVERLDOETASLSQHVFRLLDDELPEARILKLESGL 2153
QY 317 ----- 316
Db 2154 YLILAFHHTCFDAWSLKVFEBELRALYALQTKSAANLPALKAQYKEYALVHRRQLSGD 2213
QY 317 ----- 316
Db 2214 RMNLSDFWLRLKIGLEPLQLITDRPRPVQFKYDGDLDLSIELSKKETENLRGVAKRKSS 2273
QY 317 ----- 316
Db 2274 LYVWLVSVCVMLASVYANQSDSVGIPVSHRTHPQFSVIGFFVNLVLRVIDSQAICG 2333
QY 317 -----MKSIMATQLRD----- 327
Db 2334 LIRRVKMKELVDQLHDMFQFQEVTKLQVNDPDSRHPPLVQNVFNPESRANGHEHDASEDE 2393
QY 328 ----- 327
Db 2394 GSLAFNQYRPQVPDVSVAKFDLNATVTELESGLRVNFYATSLFNKSTIQGFLHTYELL 2453
QY 328 ----- 327
Db 2454 RQLSLSAAGINEDTQLSLVRPTENGDLHLPLAOSPLATTAEQKVASLNQAFERAEFLA 2513
QY 328 -----LATW- 331
Db 2514 AEKIAVQGDRAISYADLNGOANQLARYIQSVSCIGADDDGIALMLEKSIDTIIICILAIWK 2573
QY 332 ----- 331
Db 2574 AGAAYVPLDPTYPPRGVOLILEEIKAKAVLVHSSHASKCERHGAKVIAVDSPAIETAVSQ 2633
QY 332 ----- 331
Db 2634 QSAADLPTIASLGNLAVIIFTSGTSGKPKGVLBQKAVLLLRDALRERYFGRDCTKHGV 2693
QY 332 ----- 331
Db 2694 LFLSNVYVDFSVQVLVLVLSGHKLI VPPAEFVADDFYRMASHTGLSYLSGTPSLLOKI 2753
QY 332 -----VY----- 333
Db 2754 DLARLDHLQVVTAAGEBLHATQEKMRFRFNGPIYNAYVTTVTVNIIAEFTTNSIFEN 2813
QY 334 -----TTLRYRONPPCEP- 346
Db 2814 ALREVLPGTRAYVLTAALQVPVDAVGELYLAGTVTRGYLNQPLLDQVRFIENPFCKEE 2873
QY 347 -----SR-----NRTAVSE- 355
Db 2874 DIAMGRFARLYKTGDLVRSRFNRQQOQFLEYLGRGDLOIKMRGYRIEISEVQNVLTSSPG 2933
```

```
QY 356 -----FMKNTHTVL-----RNET----- 368
Db 2934 VREGAVVAKYENDTYSRTAHSLSVGYVYTTDNETVSEADILTFMKARLPTVMVPSHLCCLE 2993
QY 369 ---PYTYGTLDWSSL--- 381
Db 2994 GALPVTNGKLDVRRLEIINDSAQSSYSPRNIEAKMCRMLWESALGMERCIDDDLFK 3053
QY 382 ----- 381
Db 3054 LGDSITSLVAQHNOVGCKITVRDIPHRRTARALHDHVFVKMDSRSNVTOFRTBOGP 3113
QY 382 -----YYNETMFV----- 389
Db 3114 VIGEAFLPLIQDWFLSKALQHPMYNHTFYVRTPELDVDSLSAAVRDLOQYHDVFRMLK 3173
QY 390 ----- 389
Db 3174 REEVGFQSPAEDFSPAQLRVLVNKVDGSAVNEILDGHWGSGFNLENGPIGIGYLHG 3233
QY 390 -----ENKTASDSN 398
Db 3234 EDRSARVFWFVHMAIDTVSQILVRDLQTLYRNGSLGKSGSFRQWAEAIQNYKASDSE 3293
QY 399 KT----- 400
Db 3294 RNHWKLVMETASSISALPTSTGSRVRLSRLSPEKTASLIQGGIDRQDVSVYDSLTSV 3353
QY 401 -----TPTSPSM-----GQRTFID----- 415
Db 3354 GLALQHIAPTGPSMVTIEGHRBEVDQTLDSRTMGFTTMYFPBIPRLSTENIVQVVA 3413
QY 416 ----- 415
Db 3414 VSRPRQVARGVGYGLYGTQHPQVTVNLGQLARKQSKPKEMVLAVGDNEFEYGL 3473
QY 416 ----- 415
Db 3474 MTSPEKDRSSAVDVTAVCIDGTMIIDVDSAMSLESEQFISSIEGLNKILDGRASQ 3533
QY 416 -----PLWDYLD----- 422
Db 3534 TSREPDVQPAETTYTYFYELEPPROGPTFLPPGEGGAESYFNIVKRLRQTNMVFV 3593
QY 423 -----SLLFDEIRNFSRLSP----- 438
Db 3594 NYLHKSRLTTPLEAEMYLQVRGIQPHGFHIGWSFGILAMESRRLVASDEKIGF 3653
QY 439 -----TYVN-----LTPPEH 448
Db 3654 LGIIDTYFNVRGATRTIGLGDTEILOPIHH 3683
```

## RESULT 13

```
US-08-876-991-2
; Sequence 2, Application US/08876991
; Patent No. 5925360
; GENERAL INFORMATION:
; APPLICANT: Gregor Meyers, Tillmann R menapf,
; APPLICANT: Heinz-J rgen Thiel
; TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Organon Teknika Corporation
; ADDRESSEE: Biotechnology Research Institute
; STREET: 1330-A Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,991
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,577
; FILING DATE:
; APPLICATION NUMBER: US/08/650,584
; FILING DATE:
; APPLICATION NUMBER: US/08/469,702
; FILING DATE:
; APPLICATION NUMBER: US/08/123,596
; FILING DATE:
; APPLICATION NUMBER: 07/797,554
; FILING DATE: 22-NOV-1991
; APPLICATION NUMBER: US 07/494,991
; FILING DATE: 16-MAR-1990
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: William M. Blackstone
; REGISTRATION NUMBER: 29,772
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-876-991-2
```

```
Query Match 17.1%; Score 420; DB 2; Length 3898;
Best Local Similarity 5.0%; Pred.No. 0.0094;
Matches 188; Conservative 87; Mismatches 175; Indels 3334; Gaps 58;

QY 2 GRKEM--MYRVDVPK----- 13
Db 52 GRGDIRTLRLDLPRKDCRSNHLGPVSGIYIKPGVPYQDYTGVPVYHRAPLFEFFDEAQF 111
QY 14 -----MFVLISISFLLV-----SPINC----- 30
Db 112 CEVTKRIGRVGSDGKLYHYVCVDCILLKAKRGTPRTLKWIRNFTNCPLWVTSCDD 171
QY 31 ----- 30
Db 172 GASGSKDKPDRMNKGLKIAPREHEKDSKTPPDATIVVEGVYQIKKKGVKGNQTQD 231
QY 31 -----KVMKALY----- 38
Db 232 GLYHNKKNPPESRKKLEKALLAWAVITILLYQPVAAENITQWNLSDNGTNGIQAMVLRG 291
QY 39 ----- 38
Db 292 VNRSLHGIWPEKICKGVPTHLATDTTELKEIRGMDASERTNYTCCRLQRHEMNHKGW 351
QY 39 -----NRP----- 41
Db 352 YNIDPWIQLMNRTOQNLTEGPPDKCAVTCRYDKNDVNVVTOARNRPTTLTGCKKGKGF 411
QY 42 -----WR 43
Db 412 SFAGTVIEGCPNFVNSVEDILYGDHECGSLLODTALYLLDGMTNTIENARQGARVTSWL 471
QY 44 GLVLS----- 48
Db 472 GRQLSTAGKGLERRSKTWFGAVALSPYCNVTRKIGYIWTNNCTPACLPKNTKIIGPKGF 531
QY 49 ----- 48
Db 532 DTNAEDGKILHEMGHLSFEFLLSLVLSDFAPETASTLYLILHYAIPOSHEEPGCDTN 591
```

QY 49 -----KIGKY-----KLQKLEILRQL- 66  
Db 592 QLNLTVKLRTEDEVSSVWNIKXVCVRPDWPEYETKVALLPBAGQVVKLVLRALDLT 651  
QY 67 -----FTT-----  
Db 652 RVMSASTTAFCLIKVLGRQVVQGIWLLVTGAQRLACKEDRYAISSTNEIGLLG 711  
QY 70 -----  
Db 712 AEGLTITWKEYSHGLQDDGTAVCACTAGSKFVTALNVVSRRYLASLHRALPTSVPTEL 771  
QY 70 -----ISTKN-----  
Db 772 LFDGTNPAIEEMDDDFGLCPDTPSVIKGKXNTTLNGSAFYLVCPGIVGTGVVECTAV 831  
QY 76 -----  
Db 832 SPTTLTEVVKTRDKPPFHRVDCVTIIVEKEDLFHCKLGNWTCVKGDPVTVKGGQVK 891  
QY 76 -----  
Db 892 QCRWCGFEKPEVGLPHYPHIGKILTNETGYRVVSDTCNRCGVISTEGEHECLIGNTT 951  
QY 76 -----  
Db 952 VKVHALDERLGMPCRPKEIVSSEGVKRTSCTFNKTKLRNKYPRDSYFQQYMLKGE 1011  
QY 76 -----  
Db 1012 YOYWFNLVDTHDHYFAEFVVLVVVALLGGRVYLWLVTVIITEQLAAGLOLQGEVV 1071  
QY 76 -----VSKQPVKNLTWN-----  
Db 1072 LIGNLITHDNEVVYFLLYLVIRDEPIKKWILLFHAMTNPNVKITVALLMISGVAK 1131  
QY 88 -----TEF-----  
Db 1132 GKIDGQWQRPVTSFDIQLALAVVVVVMLLAKRDPPTPLVTVATLRTAKITNGFST 1191  
QY 91 -----  
Db 1192 DLVIATVSAALLTWTYISDYKXKTLQXLVSTVTGIFLIRVLKGIGELDLHAPTLPshr 1251  
QY 91 POYIL-----AG-----PIONSYITLWF 110  
Db 1252 PLFYLVLVISTAVTRWNLVDVAGLLQCVPITLWMTWADILTLILPTVELTKLY 1311  
QY 111 -----DFYSTQLRKPARYV-SOYNHTAKTITFRP----- 139  
Db 1312 LKEVKIGAERGWLWKTNYKRVDIYEDQTSSEGVLFPSPKQRTSAITSLWPLIKAILIS 1371  
QY 140 -----  
Db 1372 CISKWQLIYLLXIFEVSYLHKVIDEJAGTNFVSRVLAALIEVNWAFDNEEVKGLK 1431  
QY 140 -----  
Db 1432 KPELLSRVKELIHKHVRNEVVVRWFGDEIYGMFKLIGLKAATLSRNKHCMLCTVCE 1491  
QY 140 -----PPCGR-----  
Db 1492 DRWRGETCPKGRFGPPVCGMTLADPEEKHYKRIFIREDOSGGPLREEHAGYLOVKAR 1551  
QY 145 -----  
Db 1552 GOLFLRNLPLATVKVQMLLVGNLTGEIGLEHLGHWLRGPVCKVTEHERCTTSIMDKL 1611  
QY 145 -----VPSMTCLSEM-----  
Db 1612 TAPFGVMPRGTTTPRAPVRPPTSLKIRRGLETGWAYTHQGGISSVDHVTGCKDLLVCDMT 1671

QY 156 -----NVSKRNDTGE-----QGC-----GNF 171  
Db 1672 GRTRVVCQSNKMTDESEYGVKTDGCGPEGARCVCVFNPEAVNISGTKGAMVHLQKTGGEF 1731  
QY 172 T-----TNPMPF-----NVPRWN-----TKLYVG-----PTKVNVDSTI- 202  
Db 1732 TCVTASGTAPFDLKNLKGWSGLPIPEASSGRVVGRVKVGKNEDESKPTKLMSIGIQTVS 1791  
QY 203 -----  
Db 1792 ATDLTEMVKLITWNRGEFRQITLATGAGKTTELPRSVIEEIGHRRVLVLPLRAAES 1851  
QY 203 -----YFLGLTALLRYA-----  
Db 1852 VYQVMRQKPSIAFNLRIGEMKEGDMATGITVASYGYFCQMSQPKLRAAWEVSPIFLDE 1911  
QY 217 RNC-----HSFYLVNMSRN-----  
Db 1912 YHCATPEOLAIMGKIHFP-----SENLRVAMTATPAGTIVTTGQKHPIEBEPIAEVNM 1965  
QY 242 G-----TKLKNM-----RKLKR----- 255  
Db 1966 GEDLGSYLDIAGLKIPVEEMKNMLVFVPTRNMAVEAAKKAAGYNSGYYSGEDEPSN 2025  
QY 256 -----QAP-----  
Db 2026 LRVTQSQPVVVVATNAIESGVTLPDLDVVDVTGLKCEKIRLSPKMPFIVTGLKRMV 2085  
QY 259 VKEQFEKAK-----KTQST-----  
Db 2086 IGEQAQRGRGRVGRVKPGYRSQETPVGSKDYHDLLOAQRYGIEDGINITYKSFREMYD 2145  
QY 274 -----  
Db 2146 WSLYEEDSLMTQLBILNLLISELPMVAKNIMARTDHPBIQLAYNSYETQVPLFPK 2205  
QY 274 -----TTPYFSYT-----  
Db 2206 IRNGEVTTIDYNYTFLNARKLGDDVPVYVATEDEDLAVELLGLDWDPPGQGTVEAGRA 2265  
QY 282 -----  
Db 2266 LKQVVGSLSTAENALLVALGYVGYQALSKRHPVVTDIYSVEDHREDTTHLQVAPNAIK 2325  
QY 282 -----TSAALNVTTNTV-----  
Db 2326 TEGKETELKELAQGDVORCVCVAVTNVAREGIFQFMKSQALKVRETPTYKETMNTVADYVK 2385  
QY 295 -----SITTAARV 303  
Db 2386 FIEALTDKEDIKYGLWGAHTALYKSIGARLGHETAFATLVVWKLAFGGEISDHKQA 2445  
QY 304 STSTIAY-----RPD-----SSFMKSIWATQLRDLATWYTTLYR----- 339  
Db 2446 ATDLVYIINRQFPDGTETOQEGRKPFVSLVSA---LATYTKSWNNLNLKIVEPA 2502  
QY 340 -----ONP 342  
Db 2503 LATLPIAAKALKLPAPTRLESVILSTAYKTYLSTIRGKSDGLLTGTVSAAEIMSQNP 2562  
QY 343 F-----CEPSNRNTASE-----FMQN-----THVLIRNETP----- 369  
Db 2563 VSGIAMLGVGAVAHAHNAIEASEQKRTLLMKVFNFLDQAATDLVR-ESPEKIIMAL 2621  
QY 370 -----YTIYGT-----  
Db 2622 FEAVQTVGNPLRVLYHYGVFYKGEAKELAQRTAGRNLFITLIMEAVELLGVDSGKIR 2681  
QY 377 DMSS-----  
Db 2682 QLSSNYLELLYKFRDNKIKSVREIAISWAPAFSCDWTPTDDRIGLPHDNYLRVETKCP 2741  
QY 381 ----- 380



Db 2742 CGYRMKAVKNCAGELRLLEBGGSPLCRNKFGSQNRVYTKYYDDNLSEIKPVIRMEGHV 2801  
QY 381 -LYY----- 383  
Db 2802 ELYYKGTATIKLDFNNSTKVLATDKWEYDHSYLVRAKRYTGAGYRGAYLGEKPNHKLHIQ 2861  
QY 384 ----- 390  
Db 2862 RDCATITKDKVCFIKMKRGCAFTYDLSHLNLTRELIELVHKNLNLEDRIPAIVTWTMLAYT 2921  
QY 391 ----- 404  
Db 2922 FVNEDIGTIKPTGKVTPEKQBEVLQPAVVDVTVVGTSTMTTGETPTFTS 2981  
QY 405 ----- 414  
Db 2982 LGSDSKVRQVLKLGVDGQYFPGPNQORASLLEALQGVDERPSVLLIGSDKATSNRVKTAK 3041  
QY 415 ----- 438  
Db 3042 NVKIYRSRDPLELRMMKRGKILVVALSRVDTALLKEVDYKGTFLTRETLEALSGLKPKK 3101  
QY 439 ----- 438  
Db 3102 RDITKBAQWLLRLEQIBELPDWFAKBPFIPLANTKRDYHLVGDIAITIKSAKQOLGA 3161  
QY 439 ----- 438  
Db 3162 TDSKISKYGAUVYSMLSNWVQEBNKOGSLAPLPEELLQCPGQNKTHMVSAYQ 3221  
QY 439 ----- 443  
Db 3222 LAQGNWVPVSVCHVMGTIPARRTKTHPEYAVKRLVDEHKMKALCGGSLGSKHNEVI 3281  
QY 444 ----- 443  
Db 3282 GKVKYQGNLRTKMLNPGKVAQLHREGYRHNVNKTIGSVMTATGIRLEKLPVVRQTD 3341  
QY 444 ----- 450  
Db 3342 TTNFHOAIRDKIDKEENLQTPGLHKLMEVFNALKRPELEASDVANDWELERGINRKA 3401  
QY 451 ----- 450  
Db 3402 AGFFERKNIGVLDSEKNKVEEVIDSLKGNRIYYETAIPKNEKRDVNDWTAGDFVDE 3461  
QY 451 ----- 450  
Db 3522 VSFDTKANDQVTTROLELRDIQKPYFKKKWHKPIDTLTKHSEVPVISADGEVIRKG 3581  
QY 451 ----- 450  
Db 3582 QRGSGQDPTSAGNSMLNVLTMVYAFCEATGPVYKSPDRVAKIHVCGDGLITERALGEK 3641  
QY 451 ----- 450  
Db 3642 FASKGVQILYEAGKPKQITEGDKMKVAYQDDIEFCSHTFPVQVRWSDNTSSYMPGRNTTT 3701  
QY 451 ----- 450  
Db 3702 ILAKMATRLDSSGERGTIAEKAVAFSFLMYSNWPLIRICLLVLSTELQVPRPGKSTTY 3761  
QY 451 ----- 457  
Db 3762 YYEGDPI SAYKEVIGHNLFDLKRTSPFKLAKNLMSLSTLGVWTRHTSKRLQLQCVNVGTK 3821  
QY 458 NSLW 461

Db 3822 EGNW 3825  
RESULT 14  
US-09-059-853-2  
; Sequence 2, Application US/09059853  
; Patent No. 5935582  
; GENERAL INFORMATION:  
; APPLICANT: Gregor Meyers, Tillmann R menapf,  
; APPLICANT: Heinz-J rgen Thiel  
; TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Organon Teknika Corporation  
; ADDRESSEE: Biotechnology Research Institute  
; STREET: 1330-A Piccard Drive  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: U.S.A.  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/059,853  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA: 07/797,554  
; APPLICATION NUMBER:  
; FILING DATE: 22-NOV-1991  
; APPLICATION NUMBER: US 07/494,991  
; FILING DATE: 16-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: William M. Blackstone  
; REGISTRATION NUMBER: 29,772  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 258-5200  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3898 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-059-853-2  
Query Match 17.1%; Score 420; DB 2; Length 3898;  
Best Local Similarity 5.0%; Pred. No. 0.0094;  
Matches 188; Conservative 87; Mismatches 175; Indels 3334; Gaps 58;  
QY 2 GRKEM--WVRDVPK----- 13  
Db 52 GRGDIHTLRDLPRKDCRSNGHLGVPVGIYIKPGPVYQDYTGVPVYHRAPLFFDEAQF 111  
QY 14 -----MFVLISISFLIV-----SPINC----- 30  
Db 112 CEVTKRIGRTGSDGKLYHYVCVDCGICLLKLAKRGTPTLKWIRNFTNCPPLWVTSDD 171  
QY 31 ----- 30  
Db 172 GASGSKDKKPPDRMNKGLKIAPREHEKSKTKPPDATIVVEGVKYQIKKKGVKGNKNTD 231  
QY 31 -----KWSKALY----- 38  
Db 232 GLYHNKPKPPSRKKLEKALLAWAVITILLYQPVAAENITQWNLSDNGTNGIORAMVLRG 291  
QY 39 ----- 38  
Db 292 VNRSLLHGIWPEKICKGVPTHLATDTLKEIRGNMDASERTNYTCCRQORHEWKNHGWCNW 351  
QY 39 -----NRP----- 41

Db 352 YNIDPWQLMNRQTNLTEGPPDKCAVTCRYDKNDVNVVTOARNRPTTLTGCKGKNF 411  
QY 42 -----  
Db 412 SFAGTVIEGPCNVNSVEDILYDHECGSLLODTALYLLDGMWNTNIENARQARVTSWL 471  
QY 44 GLVLIS-----  
Db 472 GRQLSTAGKKLRRSKTWFGAVALSPYCNVTRKIGYIWTNNCTPACLPKNTKIIGPKF 531  
QY 49 -----  
Db 532 DTNAEDGKILHEMGHLSFLLLSLVILSDPAPETASTLYLLHYAIPQSHERPBGCDTN 591  
QY 49 -----KIGKY-----  
Db 592 QLNLTVKLRTBDVVPSSVWNIKGVCPVDPWPYETKVALLFBEAGQVILVLRALRDLT 651  
QY 67 -----ETT-----  
Db 652 RVWNSASTTAPFLICLIKVRQGVVQGIIMLLLVTAQGLACKEDYRYAISSTNEIGLLG 711  
QY 70 -----  
Db 712 AEGLTWTKEYSHGLQLDDGTAVKAVCTAGSKFVKVATLVNVSRRYLASLHKRALPTSVTFEL 771  
QY 70 -----ISTKYN-----  
Db 772 LFDGTPAIBEMDDDFGLCLPDTSPVIGKYNITLLNGSAFVLCVIGTWVVECTAV 831  
QY 76 -----  
Db 832 SPITLRTVVKTFRDKPFPHRVDCVTTIVEKEDLFHCKLGNMTCVKGDVPTYKGQVK 891  
QY 76 -----  
Db 892 QCRWGFEPKPYGLPHYPHIGKILNTNETGVVVDSTDCNRDGVVISTEGEHECLIGNTT 951  
QY 76 -----  
Db 952 VKVHALDERLGPMPCRPKFVSSEGPVKTSCTFNKYTKLKNYEPDRDSYFOQYMLKGE 1011  
QY 76 -----  
Db 1012 YOYWFNLVDTHDHYFAEFVVLVVVALLGGYVVLIVTVIILTEQLAAGLQGGQEVV 1071  
QY 76 -----VSKOPVKNLTWN-----  
Db 1072 LIGNLITHDNEVVVYFLLLVIRDEPIKKWILLPHAMTNNPVKTIITVALLMISGVAK 1131  
QY 88 -----TEP-----  
Db 1132 GKGIDGGWQRPVTSFDIQLALAVVVVVMLAKRDPPTFPLVITVATLRTAKITNGFST 1191  
QY 91 -----  
Db 1192 DLVIATVSAALLTWYISDYKKTWLOYLNVSTVGTGIFLIRVLKGTIGELDLHAPLPSHR 1251  
QY 91 POYVIL-----AG-----  
Db 1252 PLFYILVYLSTAVTRWNLVDAGLLQCVPTLLVWFTWADILILILIPYELTKLYY 1311  
QY 111 -----DFYSTQLRKPAKVY-SQYNHTAKTITFRP-----  
Db 1312 LKEVKIGAERGLWKTNYKRVNDIYEVDQTSBGVYLFPSKORTSAITSTMPLIKAILIS 1371  
QY 140 -----  
Db 1372 CISNKKQLIYLLYLIFEVSYLHKVVIDEAGTNFVSRVLAALIEVNPADNEEVKGLK 1431  
QY 140 -----

Db 1432 KFFLLSSRRVELIHKHVRNEVVVWVWFGEDEIYGMPKLIGLVKAATLSRNKHCMLCTVCE 1491  
QY 140 -----PPCGR-----  
Db 1492 DRDMRGETCPKCGRFGPPVCGMTLADPEEKHYKRIPIREDQSGGPLREEHAGYQVKAR 1551  
QY 145 -----  
Db 1552 GOLFLRNLFLVATKVKMLLVGNLGTGIDLEHLGWLRGPAVCKVTEHERCTTSMIDL 1611  
QY 145 -----  
Db 1612 TAFGVMPRGTTTPAPVRFPFTSLIKTRRGLETGWAYTHQGISSVDHVTGKDLLVCDTM 1671  
QY 156 -----NVSKRNDTGE-----QGC-----  
Db 1672 GRTRVQCNSKNMTDESEYGVKTDSCGPEGARCYFVNPENAVNISGTKGAMVHLQKTGGEF 1731  
QY 172 T-----TFNPMFF--NVPRWN-----TKLYVG-----PTKVNVDSTQI--- 202  
Db 1732 TCVTASGTPAFDLKMLKWSGLPIFEASSGRVGRVKVGNEDSKPTKLMSGIQTVSKS 1791  
QY 203 -----  
Db 1792 ATDLTEMVKKITTMNRGEPRQITLATGAGKTTELPRSVIEISGRHKRVLVLIPRAAAS 1851  
QY 203 -----YFLGLTALLRYA-----Q 216  
Db 1852 VYQYMRQKHPSTAFNLRIGEMKEGDMATGITYASYGYFCMSQPKLRAAMVVEYSFELDE 1911  
QY 217 RNCI-----HSFVLVNA MSRNL-----SENLRVAVMTATPAGITVTTTGQKHPIEBFAPEVMK 1965  
Db 1912 YHCATPEQLAIMGKIHRE-----  
QY 242 G-----TKLKNM-----RKLKRK-----  
Db 1966 GEDLGSEYLDIAGLKIPEVEKNNLVFVPTNRNNAVEAAKLRKAGVNSGYYSGEDPSN 2025  
QY 256 -----QAP-----  
Db 2026 LRVVTSQSPYVWATNAIESGVTLPLDLDVVDTGLKCEKRILSPKMPFIVTGLKRMATV 2085  
QY 259 VKEQFEKKAK-----KTOST-----  
Db 2086 IGEQAQRGRGVKPGYRYSOETPVGSKDHYDLOAQRYGIEDGINITKSFREMYD 2145  
QY 274 -----  
Db 2146 WSLYEEDSLMITQLEILNLLISEELPMVAKNIMARTDHPFIQLAYNSYETQVPVLPFK 2205  
QY 274 -----TTPVFSYT-----  
Db 2206 IRNGEVTDYDNTYFLNARKLGDVFPVYVATEDDLAVELLGLDWDPDGPGQGTVEAGRA 2265  
QY 282 -----  
Db 2266 LKQVGLSTAENALLVALFGVYQALSKRHIPVVTDIYSVEDHRLEDTHLQYAPNAIK 2325  
QY 282 -----TSAALNVTTNVTY-----  
Db 2326 TSGKETELKLAQGVQRCVEAVTNVAREGIFQMKSQLKVRTPPTYKETMNTVADVVKK 2385  
QY 295 -----SITTAARRV 303  
Db 2386 FIEALDTSKEDIKYGLWGAHTALYKISGARLGHETAPATILVVKWLAFGGESIDHIKQA 2445  
QY 304 STSTIAY-----RPD-----SSPMKSIATQRLDRLATWYTYTLRYR----- 339  
Db 2446 ATDLVWYIINRPOFGDGTETQOGRKFVASLLVSA--LATYTKSMWNNLSKIVEPA 2502  
QY 340 -----QNP 342  
Db 2503 LATLPYAAKALKLFPATPRLESVVILSTAIYKTVLSTIRCKSDGLLGTGVSAAWEIMSQNP 2562

QY 343 F-----CEPSNRRTAVSE-----PMKN-----THVLIRNETP----- 369  
Db 2563 VSVGIAVMLGVGAAHNAIEASEQKRTLLMKVFNFLDQAATDELVK-ESKEIIMAL 2621  
QY 370 -----YIYIGTL----- 376  
Db 2622 FEAVQTGNPLRLVYHLGVFYKGWEAKELAQRTAGRNLFLLIMFAVELLGVDSSEKIR 2681  
QY 377 DMSS----- 380  
Db 2682 QLSNVLLELLYKFRDNKSSVREIAISWAPAFSCDWPTDDRIGLPHDNYLRVETKCP 2741  
QY 381 ----- 380  
Db 2742 CGYRKAVKNCAGELRLLEGGSFCLCNKFGRSQNVRTKYDYDDLNLSEIKPVIRMEGHV 2801  
QY 381 -LYY----- 383  
Db 2802 ELYYKGAATIKLDPNNSKTVLATDKWEVDHSTLVRALKRYTGAGYGLGEKPNHKLHIQ 2861  
QY 384 -----NETMFE----- 390  
Db 2862 RDCATITKDKVCPIKMKRGCAFTYDLSLNLTRLIELVHKNNLEDRIPAIVTTWLTAT 2921  
QY 391 -----NKTASDSNKTPTTS----- 404  
Db 2922 FVNEDIGTIKPTGKVTPEKQBEVVLQPAVVVDDTDVAVVVGETSTWTTGTPTFTS 2981  
QY 405 -----PSMGFORFTI----- 414  
Db 2982 LGSDSKVRQVLKLGVDGQYGPNGQASLLEAIOGVDERPSVLIILGSKATSNRVKTAK 3041  
QY 415 -----DPL-----WDVDSLLFLDEIRNFSLRSP-- 438  
Db 3042 NVKIYRSRDLEREMMKRGKILVVALSRVDTALLKFDYKGTFLTRTLEALSGLKPKK 3101  
QY 439 ----- 438  
Db 3102 RDITKABQWLLRLLEDQIBELPDWPAAKEPIPLEANIKRDKYHLVGDVIATIKAKQLGA 3161  
QY 439 ----- 438  
Db 3162 TDSTKISKEGAKVYSKMLSNWVIOBENKQSLAPLFEELLOQCPGQNKTHMVSAYQ 3221  
QY 439 -----TYVNL----- 443  
Db 3222 LAQGNVVPVSVCHVFMGTIPARTKTHPYEAYVKLRELVDHMKKALCGSGSLSKHNEWVI 3281  
QY 444 ----- 443  
Db 3282 GKVKYQGNLRTKHLNLPKGVAEQHREGYRHNVYNTIGSVMTATGIRLEKLPVVRQTD 3341  
QY 444 -----TPPBHRR----- 450  
Db 3342 TTNFHOAIRDKIDKEENLQTPGLHKKLMEVFNALKRPELEASYDAVDWHEELERGNRKA 3401  
QY 451 ----- 450  
Db 3402 AGFFERKNIGEVLDSKKNVEVIDSLKKNRIRYETAIIPKNEKRDVNDWTAGDFVDE 3461  
QY 451 ----- 450  
Db 3462 KKPRVIQYPEAKTRLAITKVMYKWKQKPVVIPGYEGKTPLFQIPDKVKKEWDQFQNPVA 3521  
QY 451 ----- 450  
Db 3522 VSPDTKAWDTQVTRDLELRIDIQKFYFKKWKHKFIDTLTKHMSVFPVISADGEVYIRKG 3581  
QY 451 ----- 450  
Db 3582 QRGSGQPDTSAGNSMLNVLTMVYAFCEATGVPYKSFDRVAKIHVCGDDGFLITERALGSEK 3641

QY 451 ----- 450  
Db 3642 FASKGVQILYEAGKPOKITEGDKMKVAYQDDIEFCSHTFPVQVRWSDNTSSYMPGRNTTT 3701  
QY 451 ----- 450  
Db 3702 ILAKMTRLDDSSGERGTIAYEKAVAFSFLMYSNPLIRRICLLVLSTELQVRPGKSTTY 3761  
QY 451 -----AVNLSTS 457  
Db 3762 YVEGDPISAYKEVIGHNLFDLKRTSPKLAKNLSMTLGLVWTRHTSKRLLQDCVNVGTK 3821  
QY 458 NSLM 461  
Db 3822 EGNW 3825

RESULT 15  
US-08-476-515A-84  
; Sequence 84, Application US/08476515A  
; Patent No. 6239270  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Raab, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjalms, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; TITLE OF INVENTION: Thereof and DNA Encoding Same  
; NUMBER OF SEQUENCES: 84  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Martin Savitzky  
; STREET: Rhone-Poulenc Rorer Inc., 500 Arcola Rd.;  
; STREET: 3C43,  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Compaq PC  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: Word 7.0 (Patentin)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/476,515A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/344,836  
; FILING DATE: 23-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/SE94/00483  
; FILING DATE: 24-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9301764-8  
; FILING DATE: 24-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitzky, Martin  
; REGISTRATION NUMBER: 29,699  
; REFERENCE/DOCKET NUMBER: A1355D  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-454-3816  
; TELEFAX: 610-454-3808  
; INFORMATION FOR SEQ ID NO: 84:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4654 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-476-515A-84

Query Match 17.0%; Score 417; DB 3; Length 4654;  
Best Local Similarity 4.3%; Pred. No. 0.019;  
Matches 191; Conservative 75; Mismatches 152; Indels 4068; Gaps 63;

QY 8 VRD----- 10  
DB 130 VRDPCGADNDCCQPTCEQLTCDNGACVNTSQKCDKWDKDCSDSDSEINCTEILCHNEFS 189  
QY 11 -----VPMKFV----- 16  
DB 190 CGNGECIPRAYVCDHNDCCQSDDEHACNYPFCGQYQFTCPSGRCIYNWVWCGEDDCKD 249  
QY 17 -----LISI-----SF 22  
DB 250 NGDEDCESGPHDVHKCSPREWSPECSCISIVKVCGLDCPGRDENNTSTGKYSM 309  
QY 23 LLVSFINCK-----VMSKALYNRP--- 41  
DB 310 TLCSALNCOYCHETPYGACFCPPGYIINNSRTCVFDDCQIWIICDQKCESRGRH 369  
QY 42 ----- 41  
DB 370 LCHCEGYLBERGOYKANDSFGBASIFNSGRDLLIGDIHGRSPRILVESQNRGVAVGV 429  
QY 42 -----WRGLVLSKI----- 50  
DB 430 AFHYHLQRFVETVQNKVFSVDINGLNTQEVNLVSVETPENLAVDVWNNKIYLVETKVN 489  
QY 51 -----GKVKL----- 55  
DB 490 RIDMVLDSYRVTLITENLGHPRGIAVDPTVGYLFFSDWESLSEPKLERAFMDGNSRK 549  
QY 56 ----- 55  
DB 550 DLVTKLGPAGVTLDMISKRVYVWDSRFDYIETVYDGIQRTVHGGSLIHPFGVSL 609  
QY 56 -----DOLKLEIL----- 64  
DB 610 FEGQVFTDWTQWLVKANKFTETNPQVYQASLRPYGVTVYHSLRQPYATNCKNNGG 669  
QY 65 -----QLET-----TIST- 72  
DB 670 CEQVCVLSHRTDNDGLFRCKTGFQDLTDHERCIAVQNLIFSSQVAIRGIPFTLSQ 729  
QY 73 -----KYNVSKQPV----- 81  
DB 730 EDVMVPSGNSPFFGIDFDAQDSTIFFSDMSKMFQKIDGTGREILAAANRVENVESL 789  
QY 82 -----KNLTWNTFFQY-----ILAGPIQYSI 105  
DB 790 AFDWISKNL-----YMTDSHYKISVMRLADKTRRTVQYLNPNRSVVVHPGAYLF 841  
QY 106 TYLWFDYFSTQLRKPQYV--YQYNHTAKTI----- 135  
DB 842 FTDWF-----RPAKIMRWSGSHLLPVIINTLGMVNGLAIDWAASRLYWDAYFDK 893  
QY 136 ----- 135  
DB 894 IEHSTFDGLDRRLRHIEQWTHPGLAIFGEHLFTDWRLGAIRVRKADGEMTVIRSG 953  
QY 136 ----- 135  
DB 954 IAYILHKSVDVNIQTGSNACQHPHNGDCSHFCFPVPNFORVCGCPYGMRLASNLHC 1013  
QY 136 -----TFRRP-----PC-GR-VPSM-----TCLSEML- 155  
DB 1014 EGDPTNEPTEQCGLFSPFCNKGRCVNPYLCGDGVDCHDNSDEQLCGTLNNTCSSAFT 1073  
QY 156 -----NVSKRND----- 162  
DB 1074 CGHGECIPAHWRCKDNDVDGSDHNCNCPHAPASCLDTQYCDNHQCSKNWVCDTND 1133  
QY 163 -----TGEQCGCNFTTFNMPFNVR----- 183

Db 1134 CGDSDKNCNSTETCQPSQFNHCNHCIDLSFVCDGDKDCVDSDEVGCVLNCASQPK 1193  
QY 184 ----- 183  
DB 1194 CASGDKCIGVTNRCDGVFDCSDNSDEAGCPTRPFGMCHSDEFQCEQDGGICIPNFWECDGH 1253  
QY 184 ----- 183  
DB 1254 PDCLYSGDEHNACVPKTCPSYFHCNNGNCIHRAMLCDRDNDCGMSDEKDCPTQPRCP 1313  
QY 184 -WNTK----- 187  
DB 1314 SMOQCGLHNCVNLVSVCDGIFDCPNGTDESPLCNGNSCDFNGCTHECVQEPFGAKC 1373  
QY 188 ----- 187  
DB 1374 LCPLGFLANDSKTCEDIDECILGSCQHCYNMGRSGFRSCDTGYMLESRGTRCKVTAS 1433  
QY 188 -----LYVGTQKVVDSQT-----IYFL----- 205  
DB 1434 ESLLLVASONKIADSVTSQVHNISLVENGSIYVAVDFDSISGRIFWSDATOGKTWSA 1493  
QY 206 ----- 205  
DB 1494 FQNGTDRRVFDSIILTETIAIDWVGNLYTWTYALETIEVSKIDGSHRTVLISKNLTN 1553  
QY 206 -----GLTA----- 209  
DB 1554 PRGLALDPMEHLLFMSDMGHHPRIERASMDGSMRTVIQVDKI FWPCLGJTIDYFNRLLY 1613  
QY 210 -----LLRY----- 214  
DB 1614 FMSYLDYMDFCDYNGHHRQVIAASDLIIRHPALFLFEDSVYTTDRATRRVRANKWHG 1673  
QY 215 -----AQRNCTH-----SFY----- 224  
DB 1674 GNQSVVWNTQWPLGIVAVHPKQPNVNSPCAFSCSHLCLLSQGHFYSVCPSGWSL 1733  
QY 225 -----LVNAM----- 229  
DB 1734 SPOLLNCLRDOPFLITVRQHIIFGILNPEVKSNDAMVPIAGIONGLDVBFDAAEQYI 1793  
QY 230 -----SRNLFR----- 235  
DB 1794 WVENPGEIHRVKTGTGNTRTVFASISWVGPSMNLALDWISRLNSTYNTPTQTSIEVLTLHG 1853  
QY 236 -----VPKYI-----NGTKLK-- 246  
DB 1854 IRYKTLIANDGTALGVGPIGIVDPARGKLYWSDOGTDSGVPAKIASANMDGTSVKTL 1913  
QY 247 ----- 246  
DB 1914 FTGNLHLECYTLIDIEQKLYWAVTGRGVIERGNVDGTRMLVHLSHPWGLAVHDSFL 1973  
QY 247 ----- 246  
DB 1974 YTTDEQVEVIERVDKATANKI VLRDWNVNLQGYVYHRRNAEBSNGCSNNNACQQIC 2033  
QY 247 ----- 246  
DB 2034 LPVPGFLSCACATGFKLPNDNRSCPSYNSFIVVMSLAIRGFSLELSHSETMVPVAGQ 2093  
QY 247 -----NTMRKJK----- 253  
DB 2094 GRNALHVDVDSGFIYWCDFSSVASDNAIRRIKPDGSSLMNIVTHGIGENGVRGIAVD 2153  
QY 254 -----RK----- 258  
DB 2154 WAGNLYFTNAFVSETLIEVLTIRYRVLKVTVDMPRHI VVDPKNRYLFWADYQRP 2213  
QY 259 VKE----- 261

Db 2214 KIERSPDCTNRVTVLSEGIPTVPRGLAVDRSDGVVYVWDDSLDIARIIRINGENSEVIRY 2273  
QY 262 -----QFEKKAKTQSTT----- 274  
Db 2274 GSRVPTPYGIVTENSIIWDRNLKKIIFOASKEPENTPEPTVIRDNINWLRDVTIPDKQV 2333  
QY 275 -----TP----- 276  
Db 2334 QPRSPAENVNNPCLENNGGSHLCPALPGIHTPKDCACFTLQSDGKNCAISTENFLIPA 2393  
QY 277 -----YF-----SYTT-SA 284  
Db 2394 LNSLSRLHLDPENHSPPFOTINVERTVMSLDVSDSDRIYFTQNLASGVQISYATLS 2453  
QY 285 ALNVTNVTYSITTA-----ARRV----- 303  
Db 2454 GIHTPTVIASIGTADGIAFDWITRIIYSDYLNQMINSMAGDSNRTVIARVPKPAIRV 2513  
QY 304 -----STSTIARPDSSFWKSI MAT 323  
Db 2514 LDPCCGYLWADWDTHAKIERATLGGNFRVPIVNSSLVMPSGTL DYE----- 2561  
QY 324 QLRDLATW-----YTTLAVRON 341  
Db 2562 --EDLLYWDASLQRIERSTLTGVDRREVIVAAVHAFGLTYQYIYWTDLTYQRIYRAN 2619  
QY 342 PF----- 343  
Db 2620 KYDGSQIAMTTLNLSQPRGINTVVKNOQKQCNPCPEQFNGGCSHICAPGNAGECQCPH 2679  
QY 344 ----- 343  
Db 2680 EGNWYLANNRKHICVDNGERGASFTCSNGRCISBEWKCDNDNDGCGSDMESYCALH 2739  
QY 344 -CBPS-----RRTAVSEFMKN----- 359  
Db 2740 TCSFTAPTANGRCVQYSYRCDYNDGCGSDGAGCLFRDCNATTEFCMNRRCIPREFI 2799  
QY 360 ----- 359  
Db 2800 CNGVDNCHDNTSDEKNCPRDTCQSGYTKCHNSNICIPRYVLCGDNDGCDNSDENPTYC 2859  
QY 360 -THVLIRNE----- 367  
Db 2860 THTCSSEFQACGRCIPQHWYCDQETDCFDASDEPASCGRHSERTCLADEFKDGGRCI 2919  
QY 368 ----- 367  
Db 2920 PSEWICDGDNDGMSDEDKRHOQONQNCSDSFLCVNDRPPRRRCIPQSWCVDGVDCT 2979  
QY 368 ----- 367  
Db 2980 DGYDENQNCNTRTCSNEFTCGYGLCIPKIFRCDRHNDGDISDERGCLYQTCQONQFTC 3039  
QY 368 ----- 367  
Db 3040 QNGRCISKTFCVDEDNDCGSDGDELMLHLCHTPEPTCPPHEFKDNGRCIEMMKCNHLDD 3099  
QY 368 ----- 367  
Db 3100 CLDNDSEKGGINECHDPSISGCDHNCDTLTLSFYCSRPGYKLSMDKRTCVDDIDECTEM 3159  
QY 368 -----TPYTIYGLDMSLYY----- 383  
Db 3160 PFVCSQKCEVIGSYICKCAPYLREPDGKTCQNSNIEPYLIP-----SNRYLRNLTI 3214  
QY 384 -----NETWFEVUNKT----- 393  
Db 3215 DGYFSLILEGLDNVVALDPDRVEKRLYMDTORQVIERMFL-NKTNKETIINHLPAAE 3273  
QY 394 ----- 393  
Db 3274 SLAVDWVSRKLYWLDARLDGLFVSDLNGGHRRLAQHCVDANNVTCFDPNPRGLALHPQYG 3333

QY 394 ----- 393  
Db 3334 YLYWADGHRVAVIGRVMDGTNKSIIISTKLEWPGITTIIDYTNLLYWAHAHLGYEYSD 3393  
QY 394 ----- 393  
Db 3394 LEGHHRHTYDGAHPHPFAITIPEDTIYWTDMNTRTVBKNGKYDGSNRQTLVNTTHRPFD 3453  
QY 394 ----- 393  
Db 3454 IHVYHFRQPIVSNPCGTNNGGCSHLCLIKPGKGFTCECPDDPRTLQLSGSYCHPMCS 3513  
QY 394 -----ASDSNKTTT----- 402  
Db 3514 STQFLCANNEKCIPIWVKDQKDCSDGDELALCPQRCFLGQFQCSGNCNCTSPQTLN 3573  
QY 403 ----- 402  
Db 3574 AHQNCPDGSEDRLLCENHHCSNEWQCANRCIPESWQCDTPNDCEDNSDESSHCASR 3633  
QY 403 ----- 402  
Db 3634 TCRPGQFRANGRCIPQAMKCDVNDGCDHSDDEPIECSMAHLCDNFTFESCKTYRCI 3693  
QY 403 ----- 402  
Db 3694 PKWAVCNGVDDCRDNDSEOGCEBERTCHPVGDPRCKNHHCIPLRWQCDGQDNDGNSDEN 3753  
QY 403 ----- 402  
Db 3754 CAPRECTESEFRVNOQCIPSRWICDHYNDGNSDERCEMRTCHPEYFQCTSGHCVHS 3813  
QY 403 ----- 402  
Db 3814 ELKCDGSDCLDASDEADCTPRPDGAYCOATMFECKNHVICPPYWKCDGDDDCGDSDE 3873  
QY 403 -----TSP----- 405  
Db 3874 ELHLCLDVPNSPFRCDNNRCIYSHVCNGVDCGDTDETEBEHKRPTPKPCTEY 3933  
QY 406 ----- 407  
Db 3934 KCGNGHCIPHDNVDDADCGDMSDELGCNKGKERTCAENICEQNTQLNEGGFICSTA 3993  
QY 408 GFORTPID----- 415  
Db 3994 GFTNVFDRITSLDINECEQFGTCPOHCRNTKGSYECVACDGFMSDRPGKCAEGSS 4053  
QY 416 PL----- 417  
Db 4054 PLLLLPDNVIRKYNLSSRFSYQLDEBYIOAVDYDWDPKDIGLSVYVTVRGEGRFG 4113  
QY 418 ----- 417  
Db 4114 AIKRAYIPNPFESGRNNLVOESVDLKLKYVMQPDGIAVDWGRHYVSDVKNKRIEVALDG 4173  
QY 418 -----W----- 418  
Db 4174 RYRKMLISTDLQPAIAVNPGLMFWTDWGKPKIESAWMNGEDRNILVFDLGNPTG 4233  
QY 419 ----DYLDLSLF----- 426  
Db 4234 LSIDVLDRIYMSDFKEDVIETIKYDGTDRRVIAKEAMNPYSLDIPEDQLYMSKEKEV 4293  
QY 427 -----LDEIRNFS-----LRSTY----- 440  
Db 4294 WKONKFGQKSKETLVNPNWLTQVRIFHQLRYNKSVPNLCKQICSHLCULLRPGGSCAP 4353  
QY 441 -----VNLTP----- 446  
Db 4354 QSSSFIEGSTTECDAAIELPILNPPPCRMHGNCYFDETDLPKCKCPSGYTKYCEMAF 4413

Qy	447	-----EHRVNL-----	454
Db	4414	SKGISPGTTAVALLTILLVIVIGALAIAGFFHYRTGSLLPALPKLPSSLVKPSENG	4473
Qy	455	-----	454
Db	4474	NGVFRSGADLNMIDIGVSGFPETAIDRSMAISEDVVMEMGKQPIIPENPMYSARDSAVK	4533
Qy	455	-----	461
Db	4534	VVQPIQVTVSENVNDKNYGPINPSEIVPETNPTSPADGTQVTKW	4579

Search completed: December 7, 2004, 15:41:38  
Job time : 131 secs